KTH Computer Science
and Communication

Machine Learning Techniques with Specific Application to the Early Olfactory System

Benjamin Auffarth

Som med tillstånd av Kungliga Tekniska högskolan
framlägges till offentlig granskning för avläggande av teknologie doktorsexamen
fredagen den 16 mars 2012 kl 10:00
i sal D3, Lindstedtsvägen 5
Kungliga Tekniska högskolan, Stockholm.

TRITA-CSC-A 2012:01
ISSN-1653-5723
ISRN-KTH/CSC/A–12/01-SE

Abstract

This thesis deals with machine learning techniques for the extraction of structure and the analysis of the vertebrate olfactory pathway based on related methods. Some of its main contributions are summarized below.

We have performed a systematic investigation for classification in biomedical images with the goal of recognizing a material in these images by its texture. This investigation included (i) different measures for evaluating the importance of image descriptors (features), (ii) methods to select a feature set based on these evaluations, and (iii) classification algorithms. Image features were evaluated according to their estimated relevance for the classification task and their redundancy with other features. For this purpose, we proposed a framework for relevance and redundancy measures and, within this framework, we proposed two new measures. These were the *value difference metric* and the *fit criterion*. Both measures performed well in comparison with other previously used ones for evaluating features. We also proposed a Hopfield network as a method for feature selection, which in experiments gave one of the best results relative to other previously used approaches.

We proposed a genetic algorithm for clustering and tested it on several real–world datasets. This genetic algorithm was novel in several ways, including (i) the use of intra–cluster distance as additional optimization criterion, (ii) an annealing procedure, and (iii) adaptation of mutation rates. As opposed to many conventional clustering algorithms, our optimization framework allowed us to use different cluster validation measures including those which do not rely on cluster centroids. We demonstrated the use of the clustering algorithm experimentally with

several cluster validity measures as optimization criteria. We compared the performance of our clustering algorithm to that of the often–used fuzzy c–means algorithm on several standard machine learning datasets from the University of California/Urvine (UCI) and obtained good results.

The organization of representations in the brain has been observed at several stages of processing to spatially decompose input from the environment into features that are somehow relevant from a behavioral or perceptual standpoint. For the perception of smells, the analysis of such an organization, however, is not as straightforward because of the missing metric. Some studies report spatial clusters for several combinations of physico–chemical properties in the olfactory bulb at the level of the glomeruli. We performed a systematic study of representations based on a dataset of activity–related images comprising more than 350 odorants and covering the whole spatial array of the first synaptic level in the olfactory system. We found clustered representations for several physico–chemical properties. We compared the relevance of these properties to activations and estimated the size of the coding zones. The results confirmed and extended previous studies on olfactory coding for physico–chemical properties. Particularly of interest was the spatial progression by carbon chain that we found. We discussed our estimates of relevance and coding size in the context of processing strategies. We think that the results obtained in this study could guide the search into olfactory coding primitives and the understanding of the stimulus space.

In a second study on representations in the olfactory bulb, we grouped odorants together by perceptual categories, such as floral and fruity. By the application of the same statistical methods as in the previous study, we found clustered zones for these categories. Furthermore, we found that distances between spatial representations were related to perceptual differences in humans as reported in literature. This was possibly the first time such an analysis had been done. Apart from pointing towards a spatial decomposition by perceptual dimensions, results indicate

that distance relationships between representations could be perceptually meaningful.

In a third study, we modeled axon convergence from olfactory receptor neurons to the olfactory bulb. Sensory neurons were stimulated by a set of biologically–relevant odors, which were described by a set of physico–chemical properties that co–varied with the neural and glomerular population activity in the olfactory bulb. Convergence was mediated by the covariance between olfactory neurons. In our model, we could replicate the formation of glomeruli and concentration coding as reported in the literature, and further we found that the spatial relationships between representational zones resulting from our model correlated with reported perceptual differences between odor categories. This shows that natural statistics, including similarity of physico–chemical structure of odorants, can give rise to an ordered arrangement of representations at the olfactory bulb level where the distances between representations are perceptually relevant.

Keywords: feature selection, image features, pattern classification, relevance, redundancy, distributional similarity, divergence measure, genetic algorithms, clustering algorithms, annealing, olfactory coding, olfactory bulb, odorants, glomeruli, property–activity relationship, olfaction, plasticity, axonal guidance, odor category, perception, spatial coding, population coding, memory organization, odor quality.

Acknowledgments

My foremost gratitude goes to Anders Lansner and Tony Lindeberg, who generously took on supervision of this thesis project at an advanced stage and who provided a rich source of inspiration and motivation. Both commented on and contributed to drafts of the present thesis and also gave advice during the writing process of several articles contained therein. I want to thank Anders for introducing me to his way of modeling and Tony for introducing me to his systematic way of thinking.

Secondly, I want to express my gratitude to the people who read and commented on this thesis and helped improving it by spotting mistakes and indicating improvements. I was extremely fortunate in finding enormous support by Tony Lindeberg, Anders Lansner, Jan Pieczkowski, Bernhard Kaplan, and David Silverstein.

A thesis reflects personal growth and development, but is also the result of the working together of many factors, many of them social. I am happy to have been able to count on the support and encouragement by many people. It is not always easy to know and remember everybody and I decided on a minimal collection of people to thank personally. I like to name especially Erola Pairó, Tony Lindeberg, Hossein Farahani, Malin Sandström, Bernhard Kaplan, and Kalia Dogbevi. Many more people helped to make this world a better place for me and I would like to thank more people for their companionship and friendship, particularly my co–workers from the different labs I worked in: (in no particular order) Fabian Schreiber, Kai Du, Erik Fransèn, Tim Pearce, Marta Padilla, Alexandre Perera, Luís Fernández, Víctor Pomareda, Oriol Canela, Marc Weber, Teresa Ciurana, Javier Buceta,

Adnan Ghori, Erik Sjölund, Hassan Foroughi Asl, Jordi Martí, Ramon Roy, Kristoffer Forslund, Laura Gustafsson, Maciej Winnicki, Manuel Alonso, Miquel Pairó, Roser Castiñeira, Norman Salazar, Pierrot Bourreau, Roland Orre, David Silverstein, Anders Ledberg, Henrik Lindén, Toomas Kirt, Ronald Biro, Luis Romanidy Hurtado, and many more including all members of CB. I also want to thank my family, particularly, my parents and my sister.

Two of the presented studies were only possibly because of the tremendous work of the group around Michael Leon and Brett Johnson at the University of California at Irvine, who collected the data over the course of many years and made it publicly available. I would like to thank them. Also I feel very privileged to have met Kensaku Mori and Manuel Zarzo and I want to thank them for introducing me to the systems perspective of olfaction and the perceptual dimensions of odors. Bernhard Kaplan did proofreading and commented on drafts of all three olfaction papers. Bernhard was also often the first person to hear of my ideas and I am grateful for his constructive criticism and his cheerful attitude. Further acknowledged is work by Simon Benjaminsson in implementing a C++ library for simulation of neural networks, which was initially used for some simulations that eventually led to paper VI. Reviewers and editors at Frontiers in Systems Neuroscience were extremely helpful and friendly with comments and large and small suggestions, and patiently helped to improve the articles: Federico Bermudez-Rattoni, Stelios M. Smirnakis, Raúl G. Paredes, Raphael Pinaud, Max L. Fletcher, David C. Willhite, Milagros Gallo, and Edmund Rolls.

This work was possible because of grants from several institutions. Financial support is hereby gratefully acknowledged by the Spanish ministry of Science (MAT-2005-07244-C03-03), the university of Barcelona (inovació de docencia), the federal state government of Catalonia (formació de personal investigador, FI), and lastly, by the EU FP7 Program project NEUROCHEM under grant agreement no. 216916.

Contents

Chapter 1

General introduction

A basic problem that humans and other animals face when moving through the environment is to find structure in the world. This implies several issues including (i) detection of stimuli, (ii) discrimination of certain stimulus patterns from others, and (iii) assigning meaning to stimuli in order to generate an appropriate response and ultimately help surviving. This last point particularly implies to derive interpretations or representations of the environment and the behavioral context. In perception, peripheral systems or front–ends first filter out information. This information is continuously filtered and transmitted within a complex network architecture. How these processes work in biological systems, how they relate and fit together, is a principal question in systems neuroscience and computational neuroscience. How these processes should or can work is a question about machine learning. In this treatise, the general problem of finding structure in the world is approached from the point of view of machine learning and systems neuroscience.

From an evolutionary perspective, the sensors should extract information from the world which is relevant for the organism. Relevance can be understood in terms of the emotional valence, the hedonic value, or the afforded behavioral importance of the stimulus. In classification tasks as performed by statistical or algorithmic approaches by machines, relevance can be understood in terms of helping to disambiguate between objects.

Some parts of the world are more important than others and need more resolution. Sensors therefore should focus the attention on particular objects that match

these relevance criteria or transform the representations in a way as to emphasize informational structure.

Machine learning techniques can be seen from different perspectives:

- They can be seen in an *explanatory framework*, where they are applied in order to establish relationships. In our context, they can help to elucidate workings in biological organisms.
- They help to understand learning and principles of information processing. This is the *computational framework*. This can be done empirically and mathematically.
- They can be applied within a *modeling framework* to predict the behavior of a system.

In the work that led to this thesis, we took all three aforementioned approaches.

Generally, more and more datasets become available in neuroscience, and this offers great opportunities to researchers — especially those without access to a wetlab, or those in their early career stages — to come up with interesting and meaningful ways to analyze them.

This chapter briefly summarizes work and contributions of the author. A more thorough summary will follow in dedicated chapters. A list of papers on which this thesis is based is given in section 1.3. The different papers will be referred to by numbers according to chronological order.

1.1 Brief summary of work

In brief, papers I, II, and III are empirical studies of machine learning algorithms with some theoretical parts. First, several methods of analysis are developed and evaluated. Informational filters are proposed with the purpose of the extraction and selection of information. Then, a clustering algorithm is discussed. In papers IV, V, and VI we applied statistical and machine learning techniques to a dataset of rat physiological responses for definitions of a neuroscience problem within the aforementioned explanatory framework. This helped us to draw conclusions about representations of smell in the olfactory bulb of the rat. In paper VI, we formulated

a model of a biological system, an early part of the vertebrate olfactory system, and generated predictions that give plausibility to certain hypotheses about the organization of the nervous system.

Machine learning techniques for classification and feature selection were investigated in papers **I** and **II**. The idea was to improve a classification of tissue in computer–tomography images. For this purpose, we worked on increasing the quality of the feature set for the classification, which as we showed could help a lot. This was done within the minimum redundancy–maximum relevance framework where the goal is to optimize the trade–off between redundancy and relevance of your selected feature set. The contribution of **paper I** was (i) to compare several algorithms for feature selection and (ii) to develop a new one which used a Hopfield network. In **paper II** we developed, tested, and compared measures for evaluating redundancy and relevance. We proposed two new measures which were shown in the comparison to work well in the evaluation of relevance and redundancy.

A clustering technique based on genetic algorithms was introduced in **paper III**. It used annealing and a local fitness and offered the possibility to use many different distance measures. The algorithm makes use of cluster validity functions to create permutations as well as cross-overs of partitions with best single clusters, and it maintains gene expression likelihoods based on survival rates. We applied the algorithm to real–world data sets and used several objective criterion functions based on entropy, and inter–cluster and intra–cluster distances. The main contributions were (i) to present a genetic algorithm that is fast and able to converge on meaningful clusters, and (ii) to define several multi-variate cluster validity criteria, one based on entropy, and two based on the Mahalanobis distance and cluster–compactness.

The early vertebrate olfactory system was analyzed in the other three papers. Analyses in papers **IV** and **V** was performed on a publicly available dataset of physiological responses in rats to a set of several hundred odorants. The topic of these papers are the representations of odorant properties in the rat brain. In **paper IV**, representations of physico–chemical properties were investigated using a machine–learning strategy. To our knowledge, this was the first reported study on a large dataset (362 odorants), as opposed to many studies on odorant subgroups (typically about ten different). We could replicate results from the literature, which

were mostly based on restricted datasets, and we could extend these results. In this paper, we showed representational areas for different physico–chemical properties including different molecular bonds and functional groups. Further, we confirmed zones of chemical progression for molecular length and carbon number. Our definition and computation of relevance of physico–chemical properties was novel and permitted us to systematically compare several molecular properties by their the importance for glomerular coding. We also compared the size of the representational zones and we are also not aware of other studies that systematically evaluated sizes of coding zones for molecular properties.

In **paper V**, we performed another analysis on the dataset of rat physiological responses to odorants. By statistical analysis, we found and continuous zones of representations for perceptual odor categories. This was a novel result and we compare related articles in section 6.5 of this thesis. Furthermore, in this article, we investigated hypotheses about coding at the olfactory bulb level. We found that distances between spatial codes were more meaningful in terms of perceptual differences, than were differences between population codes.

Axon guidance in the early vertebrate olfactory system was modeled in the study leading to **paper VI**. The model predicted the formation of glomeruli, a concentration–dependent spatial increase of glomerular recruitment of coding zones, and principles of spatial relationships between the perceptual representations. This model helped us to draw conclusions about the organization and the statistics of the stimulus space and the organization of olfactory representations. The basic result is that the described spatial arrangement could partly be explained by stimulus statistics. This had not been shown before in such a form.

1.2 Summary of author's contributions

For papers I and II (Auffarth et al., 2008, 2010), the idea was to build a classification pipeline for CT images. This pipeline consisted of feature selection, classification, and refinement. I felt that some of the methods could be improved, so I started researching different filters, and feature selection schemes with respect to their effectiveness within a classification task. I implemented the methods, did the

programming for the different experiments, and then I monitored their execution, which took several weeks to compute on a desktop PC. I analyzed the results and wrote the papers. Maite Lopez–Sanchez was coordinating and streamlining these efforts and helped a lot to improve the writing. Jesus Cerquides brought in ideas for analyses and filters.

As for paper III (Auffarth, 2010), I had several ideas for a method on clustering. I developed the method, implemented it in MATLAB, conducted the experiments, and wrote the paper.

Work that led to paper IV (Auffarth et al., 2011b) started out from results of different statistical analyses, which I had performed during 2008. Agustin Gutierrez and I defined the scope of the paper together based on collected results. We agreed to use a support–vector machine as a way to do complex correlation analysis. Computations were done on a linux cluster (beowulf), which I had assembled during the summer of 2008, consisting of thirty–two processors. I compiled the results and wrote the paper. Agustin Gutierrez commented on drafts during the writing of the original conference paper that the journal paper was based on. Santiago Marco approved the final draft version that was submitted to Frontiers of Systems Neuroscience.

For paper V (Auffarth et al., 2011a), I conceived the ideas, performed the analysis, and wrote the paper. Santiago Marco approved a preliminary version of the article.

For paper VI (Auffarth et al., 2011c), Anders Lansner had a model of organizing neuronal connections based on correlations of activities. We reframed this into a mechanism of synaptic convergence and axonal growth. Together, we came to understand that this could lead to the formation of glomerular structures. I had the idea of using this framework for a model of topographical memory organization. I programmed a workflow for modeling within an MPI–based C++ neural network library, which was written by Simon Benjaminsson, and which I extended. I wrote MATLAB scripts for the analysis of results. I later re–implemented the model and performed analysis in MATLAB. I wrote the paper with feedback from Bernhard Kaplan and Anders Lansner. Bernhard Kaplan helped writing some parts and created several figures which show results and illustrate the model.

1.3 Publications

1.3.1 Papers included in this thesis

I: Benjamin Auffarth, Maite López-Sánchez, Jesús Cerquides. "Hopfield networks in relevance and redundancy feature selection applied to classification of biomedical high–resolution micro–CT images." P. Perner (Ed.), *Advances in Data Mining: Medical Applications, E-Commerce, Marketing, and Theoretical Aspects. Lecture Notes in Artificial Intelligence (LNAI) 5077*, Springer-Verlag, Heidelberg, 2008, pp 16-31.

II: Benjamin Auffarth, Maite López-Sánchez, Jesús Cerquides. "Comparison of relevance and redundancy measures for feature selection." P. Perner (Ed.), *Advances in Data Mining – Applications in Medicine, Web Mining, Marketing, Image and Signal Mining, Lecture Notes in Artificial Intelligence (LNAI) 6171*, Springer-Verlag, Heidelberg, 2010, pp. 248-262.

III: Benjamin Auffarth. "A genetic algorithm for clustering with biased mutation operator and local cross–over." *World Congress on Computational Intelligence (WCCI)*. IEEE Congress on Evolutionary Computation. Barcelona, Spain, July 18–23, 2010.

IV: Benjamin Auffarth, Agustín Gutierrez, Santiago Marco. "Statistical analysis of coding for molecular properties in the olfactory bulb." *Frontiers in Systems Neuroscience*, 5, 62. 2011.

V: Benjamin Auffarth, Agustín Gutierrez, Santiago Marco. "Continuous spatial representations in the olfactory bulb may reflect perceptual categories." *Frontiers in Systems Neuroscience*, 5, 1-14. 2011.

VI: Benjamin Auffarth, Bernhard Kaplan, Anders Lansner. "Map formation in the olfactory bulb by axon guidance of olfactory neurons." *Frontiers in Systems Neuroscience*, 5, 1-16. 2011.

1.3.2 Papers not included in this thesis

The following publications were produced, mostly during the PhD period 2008–2011, but were not included in the present thesis.

- Benjamin Auffarth, Yasumasa Mutou, Yasuharu Kunii. "An Artificial System for Visual Perception in Autonomous Robots." *Proceedings of the ninth International Conference on Intelligent Engineering Systems*, September 2005.

- Benjamin Auffarth. "Spectral Graph Clustering." *Technical report.* Departament de Matemàtica Aplicada i Anàlisi (MAIA), Universitat de Barcelona. 2007.

- Benjamin Auffarth, Maite López-Sánchez, Jordi Campos, Anna Puig. "System for Automated Assistance in Correction of Programming Exercises." *5th International Congress of University Teaching and Innovation (CIDUI).* Lleida, Spain. July 2–4th, 2008. pp. 104 (1–9).

- Benjamin Auffarth. "How to Build a Linux Cluster for Scientific Computing." *Technical report.* Department of Electronics, Universitat de Barcelona. 2009.

- Matteo Falasconi, Agustín Gutierrez, Benjamin Auffarth, Giorgio Sberveglieri, Santiago Marco. "Cluster Analysis of the Rat Olfactory Bulb Activity in Response to Different Odorants." *Proceedings of the 13th International Symposium on Olfaction and Electronic Nose (ISOEN).* Brescia (Italy), 15–17 April 2009. ISBN: 978-0-7354-0674-2. Matteo Pardo, Giorgio Sberveglieri (eds).

- Benjamin Auffarth, Agustín Gutierrez, Santiago Marco. "Relevance and Loci of Odorant Features in the Rat Olfactory Bulb." *Proceedings of Biosignal Conference.* Valencia, Spain, 20–23 January 2010.

- Benjamin Auffarth, Anders Lansner. "Activity–Dependent Memory Organization in the Early Mammalian Olfactory Pathway for Decorrelation, Noise Reduction, and Sparseness–Enhancement." *BMC Neuroscience.* 5(1), pp. 1–7, 2011.

Part I

Finding structure in the world

Chapter 2

Brief introduction to machine learning techniques

The term *machine learning* is applied to algorithms that allow computers to acquire knowledge or behaviors based on data. The first part of this thesis deals with machine learning techniques and the aim of the present chapter is to give background for the first three papers. Papers I and II deal with feature selection and classification. Paper III deals with a clustering method based on a genetic algorithm. Readers unacquainted with the topic should find enough information in this chapter to be able to understand important concepts in the field. For in–depth treatment the reader is referred to textbooks, such as Duda et al. (2000) or Han et al. (2011).

With the evolution of information technology and the permeation of open research, many datasets have become available and amenable to systematic analysis. *Data mining* (sometimes also *knowledge discovery from data* or *KDD*) generally refers to the discovery of patterns in large datasets and is therefore, next to machine learning, the term, we will use in the present treatment to refer collectively to techniques used to computationally analyze data. Data mining algorithms can be applied in many fields and for many different kinds of data. Patterns that can be searched for include relationships between items in datasets and finding classes or concepts.

Two important methodologies in machine learning are *classification* and *clus-*

tering. Both are related in the way that they detect patterns in data (*pattern recognition*). Clustering refers to an unsupervised procedure, i.e. without training with labels, of grouping data items into classes (clusters) based on a measure of similarity between instances. Classification is a supervised procedure, in the sense that during a training phase category labels are provided. Generally speaking, classification involves building a model to assign instances into categories based on training set with known categories (labels) and then to apply this model to assign labels to new instances.

One way to understand an *item* (alternatively called *instance*) is to see it as a geometric object, a *point*, embedded in space, often a multi–dimensional space. Each of the dimensions in this space are described by what in this thesis will be variably referred to as an *attribute*, a *feature*, or a *variable*, so that you can describe an item by a vector, where each entry stands for a value along a dimension.

An attribute can be of the following nature:

- ordinal (e.g. $\{\text{small}, \text{large}\}$)
- categorical/nominal (e.g. $\{\text{male}, \text{female}\}$),
- numerical:
  - discrete (e.g. $\{1, 20\}$, $\mathbb{N}$),
  - continuous (e.g. $[1, 20]$, $\mathbb{R}$),

A special case of discretely valued attributes and one which is quite frequent is when two values exist (often written $\{0, 1\}$ or $\{-1, 1\}$). This is called a binary attribute.

In case of high–dimensional datasets it can be advantageous to reduce the dataset to fewer dimensions. This dimensionality reduction can proceed either by choosing among the set of dimensions or features, called *feature selection*, or by transforming the original data to a lower dimensional space. This latter case can referred to as *feature extraction*. In combination with clustering specifically, the combined procedure of dimensionality reduction and clustering is sometimes called *subspace clustering*.

There are many technical issues related to machine learning. The rest of this chapter will be dedicated to a brief overview of the literature in both clustering and classification. Feature selection will be introduced in the context of classification.

2.1 Clustering

Let us start with an intuition, then some basic algorithms, and then come to some problems typically encountered in *clustering*.

Intuitively, given a set of data points, which can be for example images, pixels or voxels in image stacks, or results from chemical analysis, the goal is to find groups of data points that belong together. In general, a cluster is a collection of data items which are similar to each other and are dissimilar to items in other clusters. The relation between points is given by a similarity or distance function.

A *partition* (or *partitioning*) is the separation of a set of items into clusters. Clustering algorithms try to find partitions that make the most sense in terms of given similarity measures. They are mainly unsupervised techniques[1]. The purpose of clustering is to discover and visualize structures in data sets.

In the following, some of the most prominent approaches to clustering will be outlined.

The *k–means algorithm* (Lloyd, 1982; MacQueen, 1967) is one of the most popular clustering algorithms. Given a set of points in a d–dimensional space $x_1, \dots x_n \in X$ the aim is to find k sets, $S = S_1, \dots S_k$, where $k \leq n$, so that the following objective function, the within–cluster sum of squares is minimized:

$$\text{argmin}_S \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \tag{2.1}$$

where μ_i is the center of cluster S_i.

After initialization of clusters, the algorithm proceeds over iteration of two steps, (i) an assignment step, where points are assigned to clusters, and (ii) an update step, where cluster centers are shifted to correspond to the centroids of their points. Because this algorithm uses centroids, it can be called *prototype–based clustering*.

K–means is a *hard clustering* method which means that its result are *crisp partitions*, in the sense that each point belongs exclusively to one cluster. In *fuzzy clustering*, which produces *soft partitions*, each point has a degree of membership,

[1] With few exceptions, e.g. biclustering

$u_{i,j} \in [0,1]$, where $i = 1, \ldots n$ and $j = 1, \ldots k$ to each cluster, such as in *fuzzy c–means* (Bezdek, 1981)[2]:

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\mu_k,x)}{d(\mu_j,x)}\right)^{2/(m-1)}} \tag{2.2}$$

where μ_k is the cluster center k and $d(\,\cdot\,,\,\cdot\,)$ is the Euclidean distance function, which means that the membership of each point to a cluster is inversely related to its distance from the cluster center.

The parametrization with k and the distance metric, the Euclidean distance from cluster centroids, as criterion function is sometimes problematic. Both the k–means and the fuzzy c–means algorithm detect spherical structural clusters. They employ a parametric distance measure and centroids which makes assumptions about the shape of the clusters and do not work well with skewed or non-parametric data distributions. Fuzzy c–means introduces an additional parameter, $m \geq 1$, called fuzzifier, which controls the cluster tightness or — inversely — how *fuzzy* the cluster partitions can be. Generally, the fuzzy c–means algorithm is more robust to outliers and overlap than k–means (e.g. Mingoti and Lima, 2006).

Clustering can be also be done based on graphs or connectivity. Given a distance matrix D with elements d_{ij} we could iteratively connect or merge points or clusters to a common cluster that have least distances until some stop criterion is reached, which could be, for example, the desired number of clusters, k, or the maximum distance that we want to use to connect points. Such a clustering approach is called *agglomerative*, because in the described step–wise fashion, it heaps points together. In the case, where the distance between two clusters is defined as the distance on the closest points from each other (minimum distance), this is called *single–linkage clustering* (also *nearest neighbor clustering*). It is also possible to take the maximum point distance between two clusters, in which case the linkage criterion is called *complete linkage*), or the average (mean) distance between points of the two clusters, which is called *average linkage*).

[2]For notational consistency, I switched k for the more conventional c

Importantly, the just mentioned single linkage or complete linkage clustering methods do not use centroids, such as fuzzy c–means or k–means do, and they also do not impose spherical cluster shapes. An added advantage is that they leave the choice of which distance function to apply, a fact that can be very important depending on the distribution of data.

Other distance measures have been proposed, however, they have not found a wide application for reasons of computational efficiency or robustness. For example, it has been proposed to apply the Mahalanobis distance for clustering (e.g. Liu et al., 2008). However, the Mahalanobis distance may bring with it some difficulties with regard to the computation of the inverse of the covariance matrix.

Some examples for commonly used distance functions in clustering between two points or pairs of points (x_i and x_j) are the following:

- Euclidean distance: $\sqrt{\sum_l \left(x_i^l - x_j^l\right)^2}$
- Manhattan distance: $\sum_l \left|x_i^l - x_j^l\right|$
- Mahalanobis distance: $\sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)}$, where S is the covariance matrix.

The application of these measures depends on the structure of the data and may sometimes require preprocessing, for example normalization.

Other solutions to produce nonlinear separating hyper-surfaces between clusters involve kernel density estimation, non-parametric regression, and spectral methods (Filippone et al., 2008). For example, Gaffney and Smyth (1999) proposed a clustering method for univariate data based on probability estimates from a mixture of non-parametric regression models. Wang (2003) presented a small simulation study with clustering using non-parametric kernel regression. These solutions however, are sensitive to the choice of bandwidth parameters and lack robustness for a broad set of problems.

In spectral clustering (e.g. Luxburg, 2007; Auffarth, 2007), data are pre-processed to extract eigenvalues of the Laplacian of the distance matrix before clustering is performed. These methods find little application to real-world data because of their high computational costs.

Information entropy is a measure of the uncertainty associated with a distribution. Liu et al. present a fuzzy c-means clustering approach based on Mahalanobis distances (Liu et al., 2008). The maximum entropy principle (Beni, 1994) and Renyi's entropy (Jenssen et al., 2003) have also been proposed as a distance measure between centroids and data points. Butte and Kohane (2000) compute entropy between gene pairs and used thresholding to build clusters. Jenssen et al. (2003) applied mutual information for clustering of gene data. The estimation of probability density underlying the computation of entropy is difficult, especially in high-dimensional spaces.

Most clustering algorithms rely on distances from centroids because of the simpler optimization as compared to measures that take into account complete linkage of points. In paper III, we proposed a genetic algorithm for optimization of clusters. Our algorithm is capable of optimizing any criterion function with respect to any criterion fitness function. It is fast and — depending on the criterion function — robust enough for application to real-world data.

2.1.1 Cluster validation

The objective of clustering algorithms is that data points *within* clusters are similar, while points *between* clusters are different. Clustering algorithms are unsupervised, which means that real assignments a-priori are unknown. Assignment to clusters relies on a distance measure; in the case of genetic algorithms, the criterion function of the optimization is called the *fitness function*.

As a general guideline, these measures should favor minimal differences between points within clusters (intra-cluster distance) and maximal differences between points of different clusters (inter-cluster distance).

The distance measure can in principle be any distance function or goodness-of-fit function. A function that measures validity of partitions based on the structure of data in the clusters is called an *internal cluster validity measure*. An overview over some measures especially proposed for internal cluster validation can be found in the review by Pfitzner et al. (2009). They define desiderata for internal cluster validity functions, which include

- being able to work with different distributions,

- being robust to outliers, and
- being robust or invariant to scaling.

Internal validity measures that received special mention in their article include information-theoretic measures such as Lopez and Rajski's measures (Mántaras, 1991; Rajski, 1961), and several normalized mutual information measures (Malvestuto, 1986; Strehl and Ghosh, 2003; Fred and Jain, 2003; Kvalseth, 1987).

Other popular validity measures include Davies–Bouldin (Davies and Bouldin, 1979), Calinski–Harabasz (Calinski and Harabasz, 1974), Hartigan (Hartigan, 1973), Krzanowski–Lai (Krzanowski and Lai, 1988), and Silhouette (Rousseeuw, 1987).

For validation of our clustering method we used an *external validity measure* that compares the coincidence of clusters found with our method to the correct cluster assignments. We used the Jaccard index which measure similarity of partitions. It is based on the Rand index (Rand, 1971) which compares two hard partitions R and Q of some data.

$$\text{Rand} = \frac{a+d}{a+b+c+d} \tag{2.3}$$

where,

a denotes the number of point pairs belonging to same partition in R as well as in Q.

b the number of point pairs belonging to the same cluster in R but to different in Q.

c the number of point pairs belonging to different clusters in R but to same clusters in Q.

d the number of point pairs belonging to different clusters in R and different clusters in Q.

The term d can cause problems by becoming big and thereby bias the index. The Jaccard coefficient (Jain and Dubes, 1988) leaves out d with the motivation that point pairs which are neither in the same cluster in R nor in Q are insignificant for consistency between R and Q. Denoeud et al. (Denoeud et al., 2005) showed experimentally that the Jaccard index is approximately equally efficient to other

measures based on the Rand index, while showing lower variance. Formally, the Jaccard index is defined as:

$$\text{Jaccard} = \frac{a}{a+b+c} \tag{2.4}$$

2.1.2 Genetic algorithms for clustering

Evolutionary algorithms (Fogel, 1995) are optimization algorithms that use mechanisms inspired by biological evolution such as inheritance, mutation, selection, and crossover. At each iteration, the fitness of a population of candidate solutions is computed.

Genetic algorithms (GA), proposed by John Holland (Holland, 1992), are search heuristics which mimick the process of natural evolution. They are used for optimization and search problems. Genetic algorithms belong to the class of evolutionary algorithms in that they make use of operations that come from evolutionary algorithms. They extend evolutionary algorithms by encoding candidate solutions as strings, called chromosomes). GAs have the following phases:

- Initialization: Generate an initial population of K candidates and compute fitness.
- Selection: For each generation, select μ_K candidates based on fitness to serve as parents.
- Crossover: Pair parents randomly and perform crossover to generate offspring.
- Mutation: Mutate offspring.
- Replacement: Replace parents by offspring and start over with selection.

Genetic algorithms have found broad application to a variety of optimization problems. A theoretical characterization has been difficult (however, c.f. Rabinovich and Wigderson, 1999; Doerr and Auger, 2011). For a recent survey on the application of genetic algorithms to clustering problems, see Sheikh et al. (2008).

To name a few examples, Scheunders (1997) argued, based on their experiments, that a hybrid of k–means and a genetic algorithm depends less on initial conditions. Krishna et al. (1999) used the k–means clustering algorithm as a cross-over operation. Lu et al. (2004); Maulik and Bandyopadhyay (2000) applied genetic algorithms

for selection of cluster centers and evaluated fitness by Euclidean distances to these centers. Laszlo and Mukherjee (2007) also based their algorithm on cluster centers.

2.2 Classification

As mentioned before, in the context of machine learning, classification designates the process or result of a method for assigning categories to data points. Generally, given a data point x, classification refers to a function, also called model

$$h(x) = \hat{y} \tag{2.5}$$

where $\hat{y}$ is category predicted by $h()$. It is desirable that predicted values are close to correct values (called *targets* or *labels*) over a range of points. A classifier function $h()$ should further generalize, i.e. give reasonable results even for unseen points, i.e. points that were not used in constructing the classifier.

In order to ensure generalization of the model, cross–validation can be used when optimizing parameters of the model. In cross–validation, model parameters are optimized for a subset of the available data and then the model performance is tested on a different subset. In k–fold cross–validation, points are split into k subsets (*folds*) and at each iteration, one fold is held back for testing (*test set*), while the rest of the points (*training set*) are used for optimizing parameters (*training* or *learning*). In this way, *overfitting* can be avoided, where the model describes specific noise rather than general relationships of the data.

Classification, as opposed to *regression*, implies a finite number of values that y can take. Problem classes for classification can be categorized in *multiclass* and *binary* problems, depending on the domain of y. In papers I and II, we dealt with problems that have two target classes, so–called *binary problems*.

There are many different methods for classification. In the following some classification methods relevant to explaining papers I and II will be introduced, then the issue of performance measures is discussed.

2.2.1 Boosting

Boosting (Schapire, 1990) refers to machine learning methods that construct a single so–called *strong learner* from an ensemble of so–called *weak learners*. *Adaboost* and *GentleBoost* are incremental algorithms that adapt weak learners to the data at each iteration.

In papers I and II, we applied a boosting classifier as one of the classification methods. This choice was motivated by three considerations which are true for both GentleBoost (Friedman et al., 2000) and Adaboost:

- Boosting techniques, depending on the number of iterations, can approach any training set.[3]
- Boosting is fast (depending on number of iterations and the choice of *weak learner*).
- Boosting can be used off–the–shelf without tweaking of hyper–parameters, it is well-known and broadly available in different implementations. This makes boosting techniques better understandable and applicable to a larger range of problems.

Boosting algorithms often use linear decision criteria, *decision stumps*, as weak learners.

2.2.1.1 Adaboost

Adaboost is probably the best known boosting algorithm. Adaboost was first proposed for binary classification (but extended to multi-classes) and we used it in papers I and II. You can find a description of Adaboost.M1 for binary data in Figure 2.1.

Adaboost iteratively applies weak classifiers (or one particular classifier) to data. It adapts at each iteration logarithmically the importance (weight) of each pattern depending on the classification of the pattern (correct/incorrect) in the current iteration, creating an *additive model* from weak classifiers.

[3] See section on Adaboost for more detail.

Figure 2.1: *Adaboost. This is adapted from Friedman et al. (2000) and Freund et al. (1996).*

Input: $X = \{x^i | i = 1, \dots n\}$: data
$Y = \{y^i | i = 1, \dots n\} \subset \{-1, 1\}$: binary class labels
T: number of rounds (iterations)
Data: $W = \{w^i | i = 1, \dots n\}$:Weights
f_t: classifiers (could be the same for all rounds)
for $i = 1, \dots n$ **do**
  $w^i \leftarrow 1/n$ (initialize weights)
end
for $t = 1, \dots T$ **do**
  Train classifier f_t maximizing error function $e_t = 1_{y^i \neq f_t(x^i)} \cdot W$
  Adjustment factor: $c_t \leftarrow \log \frac{1-e_t}{e_t}$
  Weight adjustment: $w^i \leftarrow w^i \exp c_t \cdot 1_{y^i \neq f_t(x^i)}$
  Normalize w: $w^i \leftarrow \frac{w^i}{\sum_{k=1}^{n} w_k}$
end
Output: Classification $h : X \to Y$ is given by $\operatorname{sign} \sum_{t=1}^{T} c_t \cdot f_t(x)$

Adaboost's meta routine is most appropriate using a classifier which is fast to compute but is not necessarily very good. This classifier, a "weak classifier" or "base classifier", is characterized as a "rough rule of thumb" (Schapire et al., 2001). The weak learners we used were decision stumps.

Adaboost is very robust to noise and Friedman noted "[Adaboost] seems immune to overfitting" (Friedman et al., 2000), however, this last point has been refuted later (cf. Rätsch et al., 2001).

2.2.1.2 GentleBoost

In paper I and II we used *GentleBoost* (also: *Gentle Adaboost*) (Friedman et al., 2000), which gives real-valued confidences instead of only $\{-1, 1\}$. The description follows in algorithm 2.2.

Figure 2.2: *GentleBoost after Friedman et al. (2000).*

Input: $X = \{x^i | i = 1, \dots n\}$: data
$Y = \{y^i | i = 1, \dots n\} \subset \{-1, 1\}$: binary class labels
T: number of rounds (iterations)
Data: weights w^i (corresponding to each point x^i)
f_t: classifiers (could be the same for all rounds)
for $i = 1, \dots n$ **do**
  $w^i \leftarrow 1/n$ (initialize weights)
end
for $t = 1, \dots T$ **do**
  Estimate $f_t(x)$ by w-weighted least square fitting
  $F(x) \leftarrow F(x) + f_t(x)$
  Weight adjustment: $w^i \leftarrow w^i \exp -y^i f_t(x^i)$
  Normalize w: $w^i \leftarrow \frac{w^i}{\sum_{k=1}^{n} w^k}$
end
Output: Classification is sign$F(x)$. The confidence is $F(x)$.

2.2.2 Performance measures

There exist many performance scores for classification results and many performance scores are highly correlated. See Caruana and Niculescu-Mizil (2004); Ferri et al. (2009) for more discussion and experimental comparisons.

A confusion matrix (as in table 2.1) allows us to cross–tabulate occurrence frequencies of data points, for example image voxels, that are classified correctly and incorrectly for a given class assignment by a particular learning method. Its analysis can help to determine the "goodness" of our learning method. In signal detection theory, true positives (TP) are equivalent to the number of *hits*, true negatives (TN) is the number of *correct rejections*, false positives (FP) equivalent to the number of *false alarms*, and false negatives (FN) to *misses* (Fawcett, 2004).

To begin with the most common measure, *accuracy* is determined as the proportion of correctly classified examples to all classified examples, i. e. $\frac{\mathrm{TN+TP}}{\mathrm{TN+FP+FN+TP}}$. *Precision* is the ratio of the correctly predicted positive (negative) cases to all predicted as positives (negatives), i. e. $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$ ($\frac{\mathrm{TN}}{\mathrm{TN+FN}}$).

Table 2.1: ***Confusion matrix in signal processing.*** *Visualization of results from a binary classification task. TN stands for true negative, FP for false positive, FN for false negative, TP for true positive.*

		predicted	
		negative	positive
actual	negative	TN	FP
	positive	FN	TP

A measure which takes into account false negatives is *recall* (also *hit rate* or *sensitivity*), defined as $\frac{\text{TP}}{\text{TP+FN}}$. A measure which takes into account false positives, the *false positive rate*, is given as $\frac{\text{FP}}{\text{FP+TN}}$.

The *f-measure* combines recall and precision in a weighted harmonic mean, written as

$$(1+\alpha)\frac{\text{precision}\times\text{recall}}{\alpha\times\text{precision}+\text{recall}},$$

where α is real and non-negative[4] (Fawcett, 2004).

More complex combinations are the *receiver operating characteristics* (ROC) which is a two-dimensional depiction of classifier performance in terms of hit rates and false alarms and the *area under the ROC curve* (called AUC). ROC curves have proved to be especially useful for domains with skewed class distributions. AUC is a portion of the area under the unit square and therefore its values will always be between 0 and 1, though — reasonably — no classifier should have an AUC less than 0.5 which is random guessing (Fawcett, 2004).

In papers I and II, we used AUC as the measure of choice for the evaluation of feature selection methods on our data. For the same dataset, Cerquides et al. (2005) had used the *overlap metric*, (OM) as the ratio of true positives (TP) to the sum of true positives, false positives, and false negatives of the respective class, i.e. $\frac{\text{TP}}{\text{TP+FP+FN}}$. The overlap metric will reach one when all points are correctly classified and will go to zero with high numbers of false positives or false negatives.

As a convention to avoid the false impression of precision, we generally round to

[4]Traditionally $\alpha = 1$ and the corresponding f-measure is called F_1

two digits after the decimal point.

2.3 Feature selection

Feature selection is still often done manually by experts. However, due to great quantities of data it is becoming increasingly automatized. A comparison of the methods over articles by different authors is difficult, because of incompatible performance indicators, often unknown significance, and the different data sets methods are applied to. Saeys et al. (2007) reviewed research in feature selection with application to biological data.

Sets of features can be evaluated by either filters, which measure statistical properties or information content, or a performance score of a classifier ("wrapper approach"). There exist many heuristics for choosing subsets of features. Two standard iterative search strategies are forward selection and backward selection. Forward selection starts from the empty set and adds at each step a feature, which gives the most performance improvement. Backward selection starts from all features, eliminating at each iteration one or several features. Forward-backward algorithms make an initial guess of a useful feature set and then refine the guess by eliminating variables and adding new ones. Filter–based feature selection schemes can be very fast and give good results that other more search–heavy methods are not guaranteed to achieve (c.f. Guyon et al., 2004).

In the context of this work, we define the feature selection task as follows: given a selection criterion (error function) $\varepsilon(\,\cdot\,)$ and an initial feature set X with m features we want to find a subset $X^* \subseteq X$ such that $|X^*| = s$ (s for number of **s**elected features) and $X^* = \arg\min_{\bar{X}\subseteq X, |\bar{X}|=s} \varepsilon(\bar{X})$.

Many approaches to feature selection are either based on ranks ("univariate filter paradigm") and thereby do not take into account relationships between features, or are wrapper approaches which require high computational costs. In the context of the classification scenario in papers I and II, we chose a filter–based method for being fast and giving good results, which other computation-heavy methods are not guaranteed to achieve (c.f. Guyon and Elisseeff 2003). Filter–based approaches have the additional advantage of providing a clearer picture of why a certain feature subset

is chosen through the use of scoring methods in which inherent characteristics of the selected set of variables is optimized. This is contrary to wrapper-based approaches which treat selection as a "black-box" optimizing the prediction ability according to a chosen classifier.

Multivariate filter-based feature selection has enjoyed increased popularity (Saeys et al., 2007). It has been shown that the best subset of features may not be the set of the best individual features (e. g. Cover, 1974; Toussaint, 1971; Vapnik, 1995). The idea behind combining redundancy and relevance information is simple: one should use the features that taken together have the highest value for prediction and not the ones which have highest prediction value on their own.

Zhou and Peng (2007); Peng et al. (2005); Ding and Peng (2005) selected features in a framework they call "min-redundancy max-relevance" (henceforth abbreviated: mRmR) that integrates relevance and redundancy information of each variable into a single scoring mechanism. Knijnenburg (2004) compared two different approaches. In a cluster–based approach where variables are first hierarchically clustered and then, from each cluster, the most relevant feature is selected. The second method was a greedy ranking–based selection, where relevance and redundancy, measured by Pearson correlation coefficients, was integrated into a single score. Biesiada and Duch (2005) presented an algorithm that includes at each step variables starting from highest relevance and excluding variables that are redundant.

In chapter 3, we will describe feature selection, including the mRmR approach in more detail. Also, the greedy heuristic with threshold as presented in Biesiada and Duch (2005) will be explained.

In the following two chapters, methods described in this chapter will be used and combined with extensions. These extensions concern measures for evaluating the relevance and redundancy of features and methods for feature selection that were presented in paper I and paper II. This will be presented in chapter 3. Chapter 4 will summarize paper III, which describes an algorithm for clustering written to address problems with traditional clustering algorithms.

Chapter 3

Feature selection by relevance and redundancy

This chapter describes work done on feature selection, published in papers I and II. These papers proposed and evaluated several measures of estimating the quality of features and methods for feature selection. In chapter 2, feature selection and the concepts of redundancy and relevance were briefly introduced. In this chapter, work on feature selection using the framework of maximum relevance and minimum redundancy will be summarized. An experimental evaluation was performed within the context of tissue classification in a biomedical dataset.

In the beginning of this chapter, several filters for computing redundancy and relevance will be presented, where the emphasis is put on non–parametric filters that are low on computational costs. More specifically, we will summarize two measures for the evaluation of relevance and redundancy that were presented as novelties in paper II. After this, several algorithmic methods are explained for making use of the relevance and redundancy information. A Hopfield network is explained that was proposed and presented in paper I as one such method. In the end, the results from both papers I and II will be summarized together concerning comparison of feature selection methods and feature evaluation measures.

As stated before, this treatment focuses on filter–based feature selection where relevance and redundancy measures are integrated. To recapitulate from chapter 2,

relevance is the measure of "goodness" of the projection from individual features to labels, and redundancy is a measure of how similar features are between themselves. As such, both redundancy and relevance measures fall into the class of measures for *statistical dependence* or *distributional similarity*. The idea behind combining redundancy and relevance information is simple: you should take the features that together have the highest value for prediction and not the ones which alone have highest prediction value. It has been shown that the best subset of features may not be the set of the best individual features (compare Toussaint, 1971; Cover, 1974; Vapnik, 1995).

Several criteria for relevance and redundancy that were used in both papers will be summarized in the following. Two measures are based on mutual information, others are based on statistical tests and probability distributions. We also used correlation coefficients. Due to the uncertainty of the true distribution underlying data, we preferred non–parametric measures. Generally, non–parametric tests have less statistical power (i.e. the probability that they reject the null hypothesis is smaller) but are more robust to outliers than parametric tests. Of these measures, the value difference metric and the fit criterion were presented in paper II as measures for relevance and redundancy evaluation.

After looking at different measures for evaluating relevance and redundancy of a feature set, the next question is how to combine these two measures for each feature and how to rank features, and then — in the next step — how to select features based on combined information. For this purpose, different selection schemes will be discussed and — as mentioned before — a selection scheme for feature selection, based on a Hopfield network, will be explained. In the end, experimental comparisons of feature selection based on the different measures and selection schemes will be summarized.

3.1 Relevance criteria

Relevance criteria determine how well a variable discriminates between target classes. They are a measure between a feature and the class, i.e.

$$Rel(X, Y) \equiv \text{ how useful is } X \text{ for predicting } Y. \tag{3.1}$$

The relevance criteria that we used in experiments and will discuss in this chapter are:

- symmetric uncertainty (SU)
- Spearman's rank correlation coefficient (CC)
- value difference metric (VDM)
- fit criterion (FC)

Of these, symmetric uncertainty was used before as a relevance criterion (for example by Biesiada and Duch, 2005). As for Spearman correlations, we did not find a prior publication that refers to it as a relevance criterion. However, we thought it might be better to use a non–parametric measure instead of relying on linear correlations (specifically Pearson correlation), which have been used before as relevance measure (e.g. Yu and Liu, 2003). In paper II, we showed how a measure of probability difference, presented before as the *value difference metric* (Stanfill and Waltz, 1986), can be adapted as a relevance criterion. In the same paper, we proposed another measure, which we called *fit criterion*, with which we tried to find a simple generalization from the margin concept.

3.1.1 Symmetric uncertainty

Symmetric uncertainty (henceforth abbreviated *SU*) is a symmetric information-theoretic measure of similarity between two distributions X_1 and X_2. SU is defined as:

$$SU(X_1, X_2) = 2\frac{I(X_1; X_2)}{H(X_1) + H(X_2)}, \tag{3.2}$$

where $H(X)$ is the *Shannon entropy* defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{3.3}$$

and $I(X_1, X_2)$ is the *mutual information* (sometimes also: mutual dependence, henceforth abbreviated *MU*):

$$I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \tag{3.4}$$

The joint entropy $H(X_1, X_2)$ between two distributions X_1 and X_2 is

$$H(X_1, X_2) = -\sum_{x,y} p_{x \in X_1, y \in X_2} \log_2(p_{x,y}). \tag{3.5}$$

Mutual information intuitively means the amount of information that knowing either variables provides about the other. SU is a symmetric and scaled variant. It obtains a minimum of zero when the variable are independent and a maximum value of

$$SU_{\max} = 2\frac{\min\left(H(X), H(Y)\right)}{H(X) + H(Y)}, \tag{3.6}$$

when the two variables are completely mutually redundant.

We computed $SU(X, Y)$, where X is a feature and Y is a label, normalized to range $[0, 1]$ by equation 3.6.

Since X is a continuous random variable, computing SU requires the estimation of its density distribution, which is notoriously difficult. Many simplifications of MU have been proposed (e.g. Torkkola, 2003). After some preliminary studies of density estimation, some of which are explained in section 3.3, we decided eventually to apply histogram discretization with 100 equally–spaced bins.

3.1.2 Spearman's rank correlation coefficient

Spearman's rank correlation coefficient (henceforth abbreviated *CC*), denoted as ρ, is a non–parametric test of correlation, assessing how well an arbitrary monotonic function could approximate the relationship of two variables. Unlike the Pearson product–moment correlation coefficient, it does neither make assumptions of a linear relationship between two variables nor about the frequency distribution of variables (Noether, 1991).

In the calculation of the Spearman rank correlation coefficient, variables (given as vectors) are first converted to ranks. Ranking means that values are sorted in descending order and the ranks are the positions of each value in this ordering. In the case of tied ranks, each tied value shares a rank average. In a simple example, having three values $95, 13.3, 1912$, we sort starting from the highest value, which gives an ordering of $1912, 95, 13.3$ and the ranks are $2, 3, 1$.

Given the differences of ranks d between two features, the correlation coefficient ρ, is given by

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \tag{3.7}$$

where d_i is the difference between each rank of corresponding values of the two vectors, and n is the number of value pairs.

Continuing our simple example from above, given a second variable with values $1, 2, 9$, ranked as $3, 2, 1$, we compute

$$\rho = 1 - \frac{6 \times \left((2-3)^2 + (3-2)^2 + (1-1)^2\right)}{3(3^2 - 1)} = 1 - \frac{12}{3(9-1)} = \frac{1}{2}. \tag{3.8}$$

As a relevance criterion, we are interested in the strength of correlation rather than in the direction of the correlation, so we take absolute values of ρ. As our relevance criterion CC we write:

$$\mathrm{CC}(X, Y) = \left| 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \right|, \tag{3.9}$$

where d is the vector of rank differences between X and Y.

We used an open–source MATLAB implementation (c.f. Schloegl, 2002) for computing the Spearman rank correlation coefficients.

3.1.3 Value difference metric

We also wanted to find a criterion that would determine overlaps of probability distributions of classes and features. This was presented in paper II as a new measure for both relevance and redundancy.

We will refer to $\Pr(X)$ as the probability function of variable X, $\Pr(X \big| Y)$ as the probability function of X given Y, and $\Pr(X = x)$ as the probability density of X at point x. We define a simple, continuous, monotonic function that measures the overlap between two variables X_1 and X_2:

$$\left(\int \left| \Pr(X_1 = x) - \Pr(X_2 = x) \right|^p \mathrm{d}\, x \right)^{1/p}, \tag{3.10}$$

where p is a parameter. Given that $\int \Pr(x)\,\mathrm{d}\,x = 1$, total divergence would give the sum

$$\int \Pr(X_1 = x)\,\mathrm{d}\,x + \int \Pr(X_2 = x)\,\mathrm{d}\,x = 2. \tag{3.11}$$

In order to have a range between 0 and 1 we subtracted 1. We chose $p = 1$, which has been used earlier (Stanfill and Waltz, 1986; Wilson and Martinez, 1997; Payne and Edwards, 1998) under the name *value difference metric* as distance measure. This gives a very intuitive, vertical distance between the probability mass functions.

$$\mathrm{VDM}(X_1, X_2) = \int |\Pr(X_1 = x) - \Pr(X_2 = x)|\,\mathrm{d}\,x - 1. \tag{3.12}$$

For discrete values this is easier to compute: we write $\mathrm{B}_i(X) = \frac{\sharp(e_i \leq X < e_{i+1})}{|X|}$ as the bin count at bin i given edges $\{e_1, \ldots e_k\}$. An estimation of the probability density at bin i is then given by $\hat{\Pr}_i(X) = \frac{B_i+1}{\sum^k B_i+k}$[1].

Given bins spaced equally for X_1 and X_2, subtracting normalized bin counts $\hat{\Pr}()$[2], summing over k, and subtracting 1, we obtain:

$$\sum_k \left|\hat{\Pr}_k(X_1) - \hat{\Pr}_k(X_2)\right| - 1 \tag{3.13}$$

Our VDM relevance measure is based on the idea that *conditional* distributions of variable X with respect to different classes c,

$$\Pr(X_1 = x|Y = c_1), \Pr(X_1 = x|Y = c_2), \ldots \Pr(X_1 = x|Y = c_d)$$

should be distinct from each other. We define *VDM relevance* of a feature X and matched target label Y with two classes c_1 and c_2 as:

$$\mathrm{VDM}(X, Y) = \int |\Pr(X = x|Y = c_1) - \Pr(X = x|Y = c_2)|\,\mathrm{d}\,x - 1. \tag{3.14}$$

This measures the distinctness of feature X with respect to the two classes.

An extension of VDM to more than two classes is possible, however, was not used in the reported studies. One possibility is to compare the difference of the

[1] Compare section 3.3.

[2] $\hat{\Pr}$ is the estimated probability density, in this case, by a histogram.

density of a variable $\Pr(X_i|Y = c_i)$ to that of all other variables $\Pr(X_i|Y \neq c_i)$. For d classes VDM relevance can be defined as:

$$\mathrm{VDM}(X, Y) = \frac{1}{d} \sum_{i=1}^{d} \int |\Pr(X = x|Y = c_i) - \Pr(X = x|Y \neq c_i)| \, \mathrm{d}\, x - 1. \quad (3.15)$$

3.1.4 Fit criterion

Figure 3.1: ***Separating two distributions.*** *Finding a decision boundary between two distributions: a decision problem. Adapted from Duda et al. (2000).*

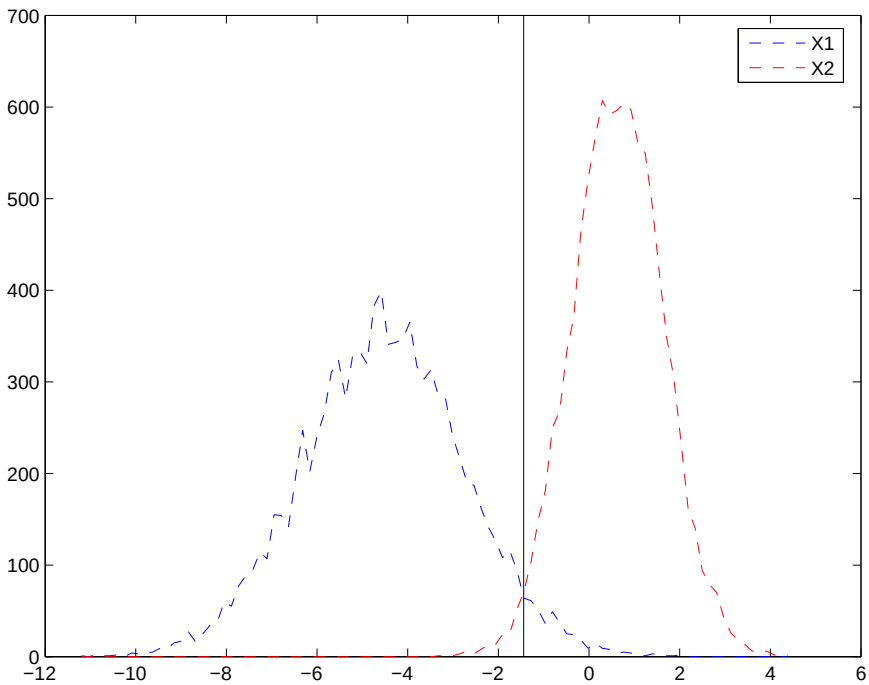


Figure 3.1 shows two (random) normal distributions which we will refer to as X_1 and X_2. No single decision boundary (threshold) on the abscissa will serve to discriminate perfectly between the two distributions. Of all thresholds the line

marks one that will lead to the best discrimination (lowest number of errors). This example is inspired from Duda et al. (2000).

The question is then: how to find a good cut between the two distributions? A sensible beginning would be to start from the centers of gravity of each class, then find a cut somewhere between the two classes. Obviously, the extent (spread) of each distribution plays an important role in finding such a cut.

For a given point x, a criterion to evaluate its typicality with respect to a distribution X_1 could be defined as the distance of the point to the center of the distribution, μ_{X_1}, in terms of the variance of the distribution σ_{X_1}.

$$\frac{|x - \mu_{X_1}|}{\sigma_{X_1}}, \tag{3.16}$$

where μX is a center of the distribution (as given, for example, by the mean [3]) and σ denotes some measure of statistical dispersion, (for example, the standard deviation).

A decision criterion as to whether a point x belongs to distribution X_1 or to distribution X_2 could be then this:

$$\text{FCP}(x, X_1, X_2) = \begin{cases} 1 & \text{if } \frac{|x-\mu_{X_1}|}{\sigma_{X_1}} < \frac{|x-\mu_{X_2}|}{\sigma_{X_2}} \\ 2 & \text{if } \frac{|x-\mu_{X_1}|}{\sigma_{X_1}} > \frac{|x-\mu_{X_2}|}{\sigma_{X_2}} \end{cases} \tag{3.17}$$

Note that you need to decide what to do in the case that both distances are equal. We refer to FCP as the *fit criterion for a given point.*

More generally for k distributions and a single feature, this can be expressed as

$$\text{FCP}(x, X_1, \ldots X_k) = \text{argmin}_{i=1\ldots k} \frac{|x - \mu_{X_i}|}{\sigma_{X_i}} \tag{3.18}$$

We will now show the derivation of the decision boundary $\dot{x}$ that results from FCP

[3] As the classical center of gravity the mean could be preferable to the median.

again given two distributions X_1 and X_2. Our decision boundary $\dot{x}$ is:

$$\frac{|\mu_{X_1} - \dot{x}|}{\sigma_{X_1}} = \frac{|\mu_{X_2} - \dot{x}|}{\sigma_{X_2}} \tag{3.19}$$

This means that $\dot{x}$ has to be at equal distance to both μ_{X_1} in terms of σ_{X_1} and μ_{X_2} in terms of σ_{X_2}. We can distinguish two cases, either $\mu_{X_1} \leq \dot{x} \leq \mu_{X_2}$ or $\mu_{X_2} \leq \dot{x} \leq \mu_{X_1}$[4]. We will continue for case 1, assuming $\mu_{X_1} \leq \mu_{X_2}$.

$$\frac{\dot{x} + \mu_{X_1}}{\sigma_{X_1}} = \frac{\dot{x} - \mu_{X_2}}{\sigma_{X_2}} \tag{3.20}$$

$$\dot{x} - \mu_{X_1} = \frac{\sigma_{X_1}}{\sigma_{X_2}} (\mu_{X_2} - \dot{x}) \tag{3.21}$$

$$\dot{x} = \frac{\frac{\sigma_{X_1}}{\sigma_{X_2}} \mu_{X_2} + \mu_{X_1}}{1 + \frac{\sigma_{X_1}}{\sigma_{X_2}}} \tag{3.22}$$

These decision boundaries are ignorant of many of the characteristics of the distributions. Note that the above expression loses meaning with long tails and with n-modal distributions ($n > 1$). An advantage of this criterion in the context of classification is that it is unbiased between different distributions, because they do not take into account prior class–probabilities.

In the case of two classes c_1 and c_2, the decision whether a given point x from distribution X_1 belongs to class c_1 or c_2 is $\mathrm{FCP}(x, X_1|Y = c_1, X|Y = c_2)$ according to eq. 3.18.

For computing relevance based on the FCP, we proceed with the conditional distributions $\Pr(X_i|Y)$ and for each point $x \in X_i$ we compute the FCP, i.e. the class which point x should belong to according to eq. 3.18. This is then matched with the target labels. The percentage of correct classification by eq. 3.18 we then call the fit criterion (short *FC*), our relevance criterion. Given data $[X|Y]$ of m features $X_i = \left\{x_i^j | j = 1 \ldots n\right\} \subset \mathbb{R}$, where n is the number of points, and matching class

[4]We could be more strict and make these two cases exclusive by assuming $\sigma_{X_1}, \sigma_{X_2} > 0$ and $\mu_{X_1} \neq \mu_{X_2}$.

labels $Y = \left\{ y^j \middle| j = 1 \dots n \right\} \in C$, we define the relevance fit criterion for binary class labels in Y and some feature X_k as:

$$\mathrm{FC}(X_k, Y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\mathrm{FCP}(x_k^i, X_k | y^i = c_1, X_k | y^i = c_2) = y^i}, \tag{3.23}$$

where $1_{\mathrm{FCP}(x_k^i, X_k | y^i = c_1, X_k | y^i = c_2) = y^i}$ is an indicator function returning 1 (correct) or 0 (incorrect) depending on the correctness of the prediction by FCP.

3.2 Redundancy criteria

We want a feature set that is not redundant. This means that each feature should provide different information for predicting the class. Redundancy criteria should capture similarities of mappings from features to classes, i.e. given a predictor function $f \in F : \mathbb{R} \to C$, our intuition is that for two non–redundant features X_1 and X_2, $f(X_1)$ should be different from $f(X_2)$. Predicting classes based on each feature is most likely associated with an error (this is related to relevance) and we hope that the effects of these errors are complementary.

Given two features X_1 and X_2, and a target Y, another intuitive formulation would be[5]:

$$\mathrm{Red}(X_1, X_2, Y) \equiv \text{ given } X_1, \text{ how much of } X_2 \text{ is dispensable for predicting Y.} \tag{3.24}$$

This statement is non–symmetric, however, we thought it intuitive that redundancy should be symmetric, i.e.

$$\mathrm{Red}(X_1, X_2, Y) = \mathrm{Red}(X_2, X_1, Y), \tag{3.25}$$

[5] Although, redundancies can be n–ary relations of features, we will henceforth take redundancies to mean binary relations, i.e. between only two features. This is simpler to estimate and also how it was used in other articles that we found on relevance and redundancy.

therefore, we formulated:

$$\text{Red}(X_1, X_2, Y) \equiv \text{ how much do } X_1 \text{ and } X_2 \text{ have in common for predicting Y?} \tag{3.26}$$

We used these redundancy criteria:

- Kolmogorov–Smirnov test on class–conditional distributions (KSC),
- Kolmogorov–Smirnov test ignoring classes (KSD),
- redundancy VDM (RVDM),
- redundancy fit criterion (RFC),
- Spearman rank correlation coefficients (CC)[6],
- Jensen–Shannon divergence (JS), and
- Sign–test (ST).

Redundancy can be measured without respect to a given class or with respect to a given class. For purpose of comparison, we include two redundancy criteria that differ only in whether or not they use class information, *KSC* and *KSD*, where KSC takes into account the classes and KSD does not.

Formally the redundancy between features X_1 and X_2 given class targets $Y \in C^n = \{c_1, \ldots |C|\}^n$ can be written as

$$\text{Red}(X_1, X_2, Y) = \frac{1}{|C|} \sum_{i=1}^{|C|} \Delta \left(X_1 | Y = c_i, X_2 | Y = c_i \right), \tag{3.27}$$

where $X_1|Y = c_i$ denotes the distribution of feature 1 given class i and Δ one of the distributional similarity measures that will be defined below. There could be more advantageous ways to combine the conditional metrics than the arithmetic mean as in equation 3.27, but we intentionally chose a conservative one.

Relevance and redundancy measures are tests for the "goodness-of-fit" and, as such, we could use the same function for measuring redundancy and relevance. Given

[6]Please, note that we did not distinguish terminology for rel_{CC} and red_{CC} as it should be clear from the context, which one is meant.

a relevance measure $Rel()$, features X_1 and X_2, and targets $Y \in C^n$, we can define

$$\mathrm{Red}(X_1, X_2, Y) = \frac{1}{|C|} \sum_{i=1}^{|C|} \left(\mathrm{Rel}(X_1|Y = c_i, X_2|Y = c_i)\right). \tag{3.28}$$

3.2.1 Kolmogorov–Smirnov test

The *Kolmogorov–Smirnov test*, short *KS–test*, can help determine whether two (one-dimensional) probability distributions could be based on the same finite samples (H_0). The D statistic of the KS-test is the maximum vertical distance between the curves of data X_1 and X_2 in an empirical distribution function (MathWorks, 2007a).

We use MATLAB's function `kstest` (MathWorks, 2007c) to obtain p-values of the 2–sided KS-test as our KS redundancy.

We used two variants of KS redundancy: Biesiada and Duch (2005) used the Kolmogorov–Smirnov test as redundancy measure, however, did not specify whether they tested class–conditional distributions $p(X_i|Y = c_k)$ or features $p(X_i)$ as a total. We assume they used $\Delta\left(p(X_i), p(X_j)\right)$ and computed the KS-test on the conditional distributions and on the features as a total. The KS-test as (presumably) used by Biesiada and Duch (2005) is given as KSD, while the KS on the conditional distribution is given as KSC.

3.2.2 Redundancy VDM

VDM, introduced as relevance criterion, can be restated as a redundancy criterion. Given a target $Y \in C^N$ and features X_1 and X_2,

$$\mathrm{RVDM}(X_1, X_2, Y) = \frac{1}{|C|} \sum_{i=1}^{|C|} \sum_{k} \left| \hat{\Pr}_k(X_1|Y = c_i) - \hat{\Pr}_k(X_2|Y = c_i) \right|, \tag{3.29}$$

where $\hat{\Pr}$ is the estimated probability.

3.2.3 Redundancy fit criterion

The fit criterion gives the goodness of fit for each point with respect to a certain class. Rating correct class predictions by this criterion as hits and incorrect predictions

as misses, the sequence of class predictions from a feature X_k can be rewritten as a binary sequence of truth values $\text{hits}_{X_k} = \left\{ h\left(\text{FCP}_{X_k^j}, Y^j\right) | j = 1, \ldots N \right\} = \{\text{hit}, \text{miss}\} = \{1, 0\}^N$, where $h()$ is an indicator function. A redundancy criterion between two such binary sequences hits_{X_1} and hits_{X_2} could be the normalized sum of hits combined by binary operators $\vee$ (or) and $\wedge$ (and):

$$\text{RFC}_{X_1,X_2} = \frac{\sum^N (\text{hits}_{X_1} \wedge \text{hits}_{X_2}) \vee (\neg\text{hits}_{X_1} \wedge \neg\text{hits}_{X_2})}{N} \tag{3.30}$$

This measure, where N is the number of points, quantifies the percentage of identically classified points after a hypothetic perfect join[7]. We will refer to the fit criterion as "RFC" when it is used to measure redundancy, specifically distinguishing it from the fit criterion for relevance.

3.2.4 Spearman's Rank Correlation

Spearman's rank correlation could be used as a measure of similarity between two features as well (again, we take absolute values of ρ). Using equation 3.28 we convert relevance to redundancy:

$$\text{Red}_{\text{CC}}(X_1, X_2, Y) = \frac{1}{|C|} \sum_{i=1}^{|C|} (\rho(X_1|Y = c_i, X_2|Y = c_i)), \tag{3.31}$$

where $\rho()$ computes the Spearman correlation coefficient of its two arguments.

3.2.5 Jensen–Shannon divergence

The *Jensen–Shannon divergence* is a variation of the *Kullback–Leibler divergence* (sometimes: *information divergence*, *information gain*, *relative entropy*, henceforth short: *KL*), which is an information theoretic measure of the difference between two probability distributions P and Q. Given the probability distributions $p(i)$, $q(i)$ of

[7] There most likely exist better options, involving graded instead of binary sequences and/or quantifying only the number of incorrectly classified points, however, our solution has the advantage of being straightforward and simple.

two discrete variables, *KL* is defined as: (Kullback and Leibler, 1951; Lin, 1991)

$$D_{\mathrm{KL}}(P\|Q) = \sum_i p(i) \log \frac{p(i)}{q(i)}, \tag{3.32}$$

and for probability distributions of two continuous variables

$$D_{\mathrm{KL}}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \,\mathrm{d}\, x. \tag{3.33}$$

Note that

$$\begin{array}{rcccc} D_{\mathrm{KL}}(P\|Q) & = & -\sum_x p(x) \log q(x) & + & \sum_x p(x) \log p(x) \\ & = & H(P,Q) & - & H(P) \end{array} \tag{3.34}$$

Here $H(P)$ denotes the entropy and $H(P,Q)$ the cross-entropy. Having $H(P)$ defined above, the cross-entropy in the discrete case is

$$\mathrm{H}(p,q) = -\sum_x p(x) \log q(x). \tag{3.35}$$

The KL–divergence is non–symmetric and non–negative. The Jensen–Shannon divergence (henceforth short: *JS*) is a symmetrized and smoothed version of the Kullback–Leibler divergence (Lee et al., 1999).

$$\mathrm{JS}(P \parallel Q) = \frac{1}{2} D_{\mathrm{KL}}(P \parallel \frac{1}{2}(P+Q)) + \frac{1}{2} D_{\mathrm{KL}}(Q \parallel \frac{1}{2}(P+Q)) \tag{3.36}$$

The units of this measure are called *bits* for $\log_2$, *hartleys* (or *bans*) for $\log_{10}$, and *nat* (also: *nit*) for $\log_e$ (which is what we used for information theoretic measures).

We take the conditional distributions, i.e.

$$\mathrm{Red}'_{\mathrm{JS}}(X_1, X_2, Y) = \frac{1}{|C|} \sum_{i=1}^{|C|} (\mathrm{JS}(X_1|Y=c_i \parallel X_2|Y=c_i)) \tag{3.37}$$

Because JS is a measure of information divergence (and not similarity), we take the inverse by subtracting from the maximum over all computed divergences in order to

obtain a redundancy measure.

$$\mathrm{Red}_{\mathrm{JS}}(X_1, X_2, Y) = \left(\max_{j=1,\dots d} \max_{k=j,\dots d} \mathrm{Red}'_{\mathrm{JS}}(X_k, X_j, Y) \right) - \mathrm{Red}'_{\mathrm{JS}}(X_1, X_2, Y) \quad (3.38)$$

3.2.6 Sign–Test

The *two-sided sign–test* (henceforth short: ST) for paired-values gives the p-value for the null hypothesis, that the difference between distributions X and Y comes from a distribution where the median is zero. The sign–test is a non–parametric test that has the advantage of simplicity and generality (Noether, 1991).

We deal with a distribution $\bar{X} = X - Y$. We compute two test statistics $S_- = |\bar{X} < 0|$ and $S_+ = |\bar{X} > 0|$, discarding all observations that equal 0.

We used p-values from MATLAB's signtest function (MathWorks, 2007e) as our ST redundancy.

3.3 Density estimation and quantization

Several of the measures previously presented need discrete data as input. Approximating a probability or a density at some point x from a sample is non–trivial. After experiments with different bin sizes and subsequent smoothing by convolutions of Parzen and Gaussian windows, respectively, and using different window sizes, we reluctantly made peace with the fact that the approximation of density depends on the scale of measurement (c.f. Mandelbrot, 1967), and stayed with histograms with $k = 100$.

Computations of symmetric uncertainty using this approach did not show a significant performance increase, and no further attempts at different parametrizations were undertaken.

It was thought that mixture modeling could possibly give better results than using histograms. Given a random sample $X = (x_1, \dots x_n)$ with a continuous, univariate density f, the univariate kernel density estimator (KDE) is

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h} \right), \quad (3.39)$$

with kernel $K(\cdot)$ and bandwidth h. The bandwidth parameter h controls the smoothness/roughness of the density estimate and unfortunate selection may give under- or over-smoothing of the data.

$K(\cdot)$ includes some restrictions such as $\int K(u)du = 1$ and $\forall u : K(u) \geq 0$. We employed the Epanechikov kernel, defined as

$$K(u) = \frac{3}{4}(1 - u^2)\ 1_{(|u|\leq 1)} \tag{3.40}$$

Such computations were performed at reduced scale[8]. It was concluded that time consumption was exorbitant and non-practical in computations involving the complete feature set.

Histograms are the most common and simplest way to approximate probability distributions. Conforming to Cromwell's rule of avoiding probabilities of 1 and 0 (except for logical true and false), we apply the Laplacian *rule of succession* (c.f. Zabell, 1989) by calculating the probabilities of bin i with frequency count n_i as

$$\tilde{p}(i) = \frac{n_i + 1}{k + \sum_{j=1}^{k} n_j} \tag{3.41}$$

We made several attempts at setting bin sizes according to different ideas, including the Nyquist–Shannon Sampling theorem (Nyquist, 1924), Scott's choice, and others (c.f. Wand, 1997). We also did bin size optimization for some features (not reported in this thesis), yet with inconclusive results. In order to avoid any problems with optimization of a bandwidth or bin number we chose a rigid bin number of 100.

Linde and Lindeberg (2012) investigated the influence of the number of bins for object recognition using histogram–based image descriptors. They found that the best results are often obtained for low numbers of bins, with just two to five bins per dimension.

[8] Using Alex Ihler's KDE toolbox, see `http://ttic.uchicago.edu/~ihler/code/`.

3.4 Feature selection schemes

We will now look at three ways to combine redundancy and relevance information in order to select features. These are:

1. minimum redundancy–maximum relevance (mRmR) — combination by measure
   (a) $\frac{\text{rel}}{\text{red}}$, or
   (b) $\text{rel} - \text{red}$
2. Greedy algorithm with redundancy threshold
3. Hopfield network — integration of redundancy and relevance information in a network

3.4.1 Minimum redundancy maximum relevance feature selection

Ding and Peng (2005), Peng et al. (2005), and Zhou and Peng (2007) presented minimum redundancy–maximum relevance feature selection. The method boils down to a forward scheme[9] maximizing one out of two measures for combination of redundancy and relevance information (mutual information in both cases) by subtraction and division, respectively. These measures are:

- $\text{argmax}_i \text{rel}(i,c) - \frac{\sum_j \text{red}(i,j)}{m}$, with i and j being two features, c the paired class, and m the numbers of competing features at each step,
- $\text{argmax}_i \frac{\text{rel}(i,c)}{\frac{\sum_j \text{red}(i,j)}{m}}$.

Peng et al. (2005) used mutual information as measure for relevance and redundancy. They referred to the first measure as mutual information difference (MID), and to the second as mutual information quotient (MIQ). We will refer to these two normalization methods (dropping the reference to mutual information) as *mRmRD*

[9] Peng et al. (2005) also discussed and tested a backward scheme but it is given less importance than the forward scheme.

and *mRmRQ*, respectively. We implemented the mRmR forward search and integrated it with our redundancy and relevance methods. The algorithm works as summarized in figure 3.2. `best()` is the selection measure, i.e. either quotient or difference. Features are $X_i, i \in [1, \dots m]$.

We did not want to mix results from different discretizations, hence we used the discretization with 100 bins. Also we did not want to rely on closed–source binaries as provided in Peng's MATLAB toolbox for computing the mutual information[10]. We implemented the forward search with our redundancy and relevance methods. We trusted the measure of mutual information should be similar to SU (and, to some extent, to JS) so that some of our methods are comparable to that of Peng et al. (2005).

The algorithm works as lined out in figure 3.2. *best*() is the function that selects the best feature by either quotient or difference. Features are $X_i, i \in [1, \dots m]$.

Figure 3.2: *mRmR feature selection after Peng et al. (2005).*

Input: rel $\in \mathbb{R}^m$: relevance scores
red $\in \mathbb{R}^{m^2}$ redundancy scores
s: number of features that need to be selected (assumed $k \geq 1$)
Initialize set $D = \{X_1, \dots X_m\}$
for $i \leftarrow 1; i \leq s; i++$ **do**
  $S_i \leftarrow best(D)$
  $D \leftarrow D \setminus S(i)$
end
Output: S: s features ordered by mRmR

3.4.2 Greedy algorithm with redundancy threshold

Biesiada and Duch (2005) presented a forward–scheme that proceeds at each step including variables starting from highest relevance and excluding variables that are redundant. They used the Kolmogorov–Smirnov test (KS) for measuring redundancy and set the threshold to a p-value> 0.95. We implemented this greedy forward

[10] See `http://research.janelia.org/peng/proj/mRMR/index.htm`

heuristic, and we tried it with different combinations of redundancy and relevance criteria.

Albeit simple as an algorithm, while assuming a meaningful computation of relevancy and redundancy, the effectiveness of the resulting set of features depends on thresholds for redundancy. It does not take into account more complex redundancies between variables, which leads to the fact that the inclusion of the individually most relevant variables can lead to suboptimal results (a common finding also reported by Biesiada and Duch (2005) in the introduction to their article). Figure 3.3 summarizes the workings of the algorithm.

Figure 3.3: *Greedy feature selection algorithm with Thresholding after Biesiada and Duch (2005).*

Input: m: number of features,
rel $\in \mathbb{R}^m$: relevance scores,
red $\in \mathbb{R}^{m^2}$: redundancy scores, ϵ: threshold
Initialize sets: set $S \leftarrow \emptyset$, and $C \leftarrow \{c_1, \dots c_m\} = \{1, \dots m\}$
while $|C| > 0$ **do**
  $S \leftarrow S \cup \text{argmax}_i \text{rel}_{c_i}$
  $C \leftarrow C \setminus \{i\}$
  $C \leftarrow C \setminus \left\{j | \exists s_i \in S, \text{red}_{s_i,c_j} \geq \epsilon\right\}$
end
Output: Selected features are in S

3.4.3 Hopfield network

In this section, a feature selection scheme based on a Hopfield network will be explained. This feature selection scheme was introduced in paper I. To our knowledge an unsupervised neural network has not been used before for *feature selection* for optimization of minimal *redundancy* and maximal *relevance*, though approaches that use neural networks for *feature selection* (e.g. Yu et al., 2004) or that apply a step of *feature selection* before feeding data into neural networks exist (e.g. Valenzuela et al., 2006).

The spaces of feature combinations and the corresponding space of their energy

or error functions have numerous local minima, which iterative algorithms could have difficulties dealing with. This brought us to the idea of applying a network as a manner of partitioning the feature space and selecting from the emergent pattern within the configuration space of the network arising from connections and activations.

Drawing each feature as a node with an activation (relevance), we can connect them to form complete graphs. The redundancy measures gives us proximity matrices $D \in \mathbb{R}^{m^2}$, where m is the number of variables. Redundancy is a (inhibitory) connection between features.

3.4.3.1 Introduction

The Hopfield network (Hopfield, 1982, 1984; Tank and Hopfield, 1986) is a well-studied recurrent network. As an introduction, we we will briefly lay out the simplest version of the Hopfield network. In the next section, we will then explain the version we used. A concise introduction to Hopfield networks can be found in Hopfield (2007).

In the simplest form of the Hopfield network, we formalize the connections (having an appropriate normalization) as symmetric, real-valued connections w_{ij}, units $S_i \in [0, 1]$ and corresponding bias units I_i. The input to each unit i is

$$n_i \leftarrow \sum_j w_{ij} S_j + I_i \,, \tag{3.42}$$

where S_j is the activation of unit j.

In the classical (binary) formalization, nodes can be asynchronously (serially) updated at each time step t:

$$S_i(t+1) \leftarrow \begin{cases} 1 & \text{if } n_i(t) > 0, \\ 0 & \text{if } n_i(t) < 0, \\ S_i(t) & \text{otherwise.} \end{cases} \tag{3.43}$$

In this formalization, the bias term I_i serves in lieu of a threshold parameter.

Unit updating proceeds iteratively by randomly choosing a unit and applying equation 3.43.

Thinking of redundancies as proximity matrices $D = \mathbb{R}^{m^2}$, where m is the number of variables, the features can be imagined as nodes. Redundancy constitutes inhibitory connections between the features. The lateral connections represented by the redundancies could create an attractor network that forms basins of attractions, where redundancies are lowest and relevancies are highest. In this manner, the choice of features could come up as an emergent pattern within the configuration space of the network arising from the connections and activations. The idea of using a Hopfield network for feature selection in such a form for feature selection was that it could have the advantage of being able of generating more complex integration of redundancy and relevance and of providing insight into the number of variables without prior knowledge.

3.4.3.2 Implementation

We chose a simple implementation for continuous (graded) activation (and responses) and asynchronous updating in discrete time-steps[11].

We tried many different parameters, normalizations of activations and connections. Parameters included annealing with different rates, using, for example, Rprop (Riedmiller and Braun, 1993). However, we found that a simple configuration was most successful. Weights and activations were normalized in the range $[-1, 0]$ and $[0, 1]$ respectively, with the diagonal of the weight matrix set to 0. We set the noise parameter $u_0 = 0.015$ (Hopfield, 1984) and we fixed the adaptation rate at $\lambda = 0.1$.

The activity function of units should be a monotone increasing function that is bounded below and above. We chose the hyperbolic tangent $\tanh x = \frac{\sinh x}{\cosh x}$ as our activation function.

The update of the activation S of a neuron i at time step t is then

$$S_i(t+1) \leftarrow (1-\lambda)\, S_i(t) + \lambda \left(1 + \tanh\left(\frac{u_i}{u_0}\right)/2\right), \text{ where} \tag{3.44}$$

[11] Our attempts at converging at a good implementation were streamlined considerably by Abdi (1994).

$$u_i = \sum_j w_{ij} \times S_j(t). \tag{3.45}$$

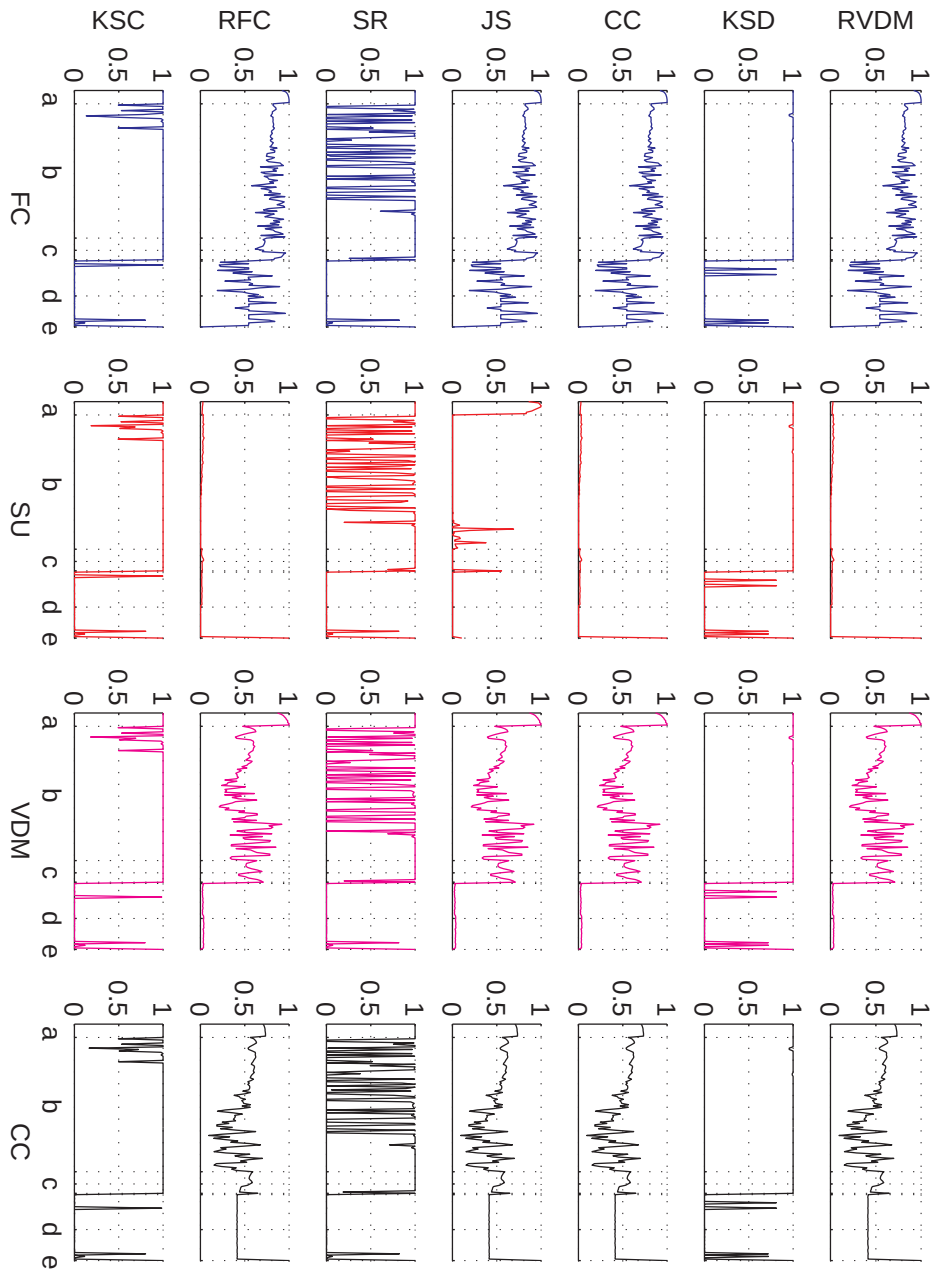
Feature selection is then — after convergence or after the course of some runs — a thresholding of the unit activations:

$$S_i \leftarrow \begin{cases} 1 & \text{if } \sum_j w_{ij} a_j > \theta, \\ 0 & \text{otherwise.} \end{cases} \tag{3.46}$$

We choose the most highly activated units in application to feature selection. Fig. 3.4 shows the activations of features after running them through a network with 1000 iterations in all combinations of relevance and redundancy measures. Each plot shows feature indices on the abscissa and activation on the ordinate. High activation means that corresponding features are assigned high importance. Rows show redundancy measures in combination with the four relevance measures. Curves differ greatly over relevance as well as redundancy measures. This means that relevance and redundancy measures each have their say in the resulting selection. From the figure, looking at importance assigned to probes, we can make predictions which relevance and redundancy combinations may work and which not. With the notable exceptions of KSC and KSD, SU seems to assign the highest importance to probes. CC relevance shows the same tendency in less degree (with exception of combinations with ST, KSD, KSC). On the other side VDM and FC do not seem to have that problem. As for redundancy measures, feature selections based on KSC and KSD make little distinctions among the real features, yet are probe–killers. More choosy is ST, which either highly rejects or highly accepts some features and yields relatively high activation for at least one probe. RVDM gives good results in combination with VDM and FC. JS performs well with VDM and — to a lesser degree — with FC.

The biological plausibility of this model is probably much better than that of supervised networks that train with back-propagation. Credence to the activating nodes gives synchronization of neural cell oscillation which correlates with visual attention (Fries et al., 1997, 2001).

Figure 3.4: ***Feature selection based on a Hopfield network.*** *This figure serves as an illustration for the distribution of weights after running the Hopfield feature selection based on different initializations. The implementation of the Hopfield network is as in paper I. In each plot, on the abscissa each point stands for one feature, the ordinate indicates network activation. Legend:* ***a*** *– Laplacians,* ***b*** *– Gabor filters,* ***c*** *– LC, TC,* ***d*** *– random,* ***e*** *– zeros.*



3.5 Evaluation

These ideas and implementations for feature selections were subjected to an exhaustive testing procedure, where all combinations of relevance and redundancy as well as selection schemes were combined and classification performance using three different classifiers was evaluated. In this section, the testing procedure and evaluation is explained. In following section, results will be summarized. The goal of these tests was to find out which selection schemes and which relevance and redundancy measures perform best. These insights were to be applied for a classification of biomaterial from a dataset, the same which was also used for the comparison of the presented methods.

3.5.1 Data set

The data consisted of slices in a 3-D volume of $423 \times 486 \times 159$ voxels, taken from CT of bones, into which a tracing material was introduced[12]. Fig. 3.5 shows the alien biomaterial marked in white on the right and organic bone material (referred to henceforth as *non–biomaterial* in order to distinguish it from the introduced biomaterial). For the classification task, the introduced biomaterial is the target class and relatively small as compared to the non–target class. The volume centers around the introduced material and hence, percentage of biomaterial is greatest in the centers (around 10 percent), becoming less towards the exteriors.

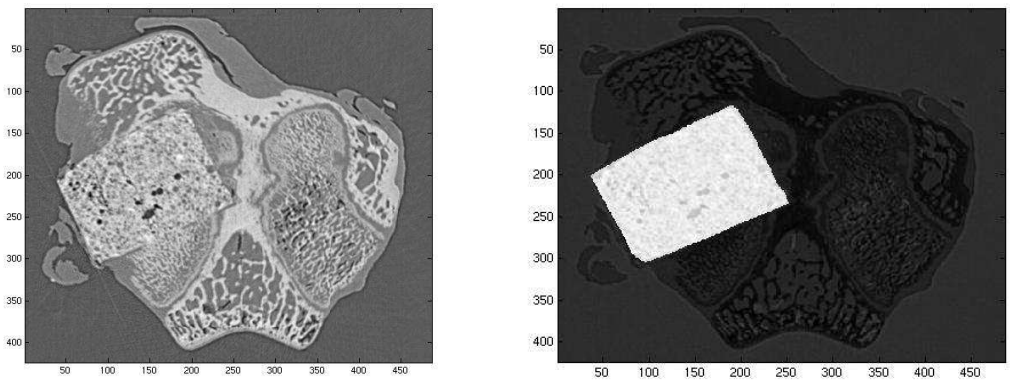
3.5.2 Feature extraction

Choosing an appropriate feature set is a classical problem in pattern recognition. Features for computational pattern recognition purposes range from local, gradient based, to global ones, which are based on the histograms of images. They can be perceptual primitives (e.g. David Marr Marr, 1982) structural descriptions (for example Irving Biederman's *Recognition by Components*; Biederman, 1987), properties which human observers perceive as salient, (e.g. Itti and Koch, 2000; Treisman and Gelade, 1980), or — in a more pragmatic vein — some features of the image,

[12]Samples available at `http://www.maia.ub.es/~maite/out-slice-250-299.arff`.

Figure 3.5: ***Dataset of biomaterial used in benchmarks of feature selection.*** *Slices from the computerized–tomography image dataset. Left: a central slice from the 3-D volume. Right: labels marked white for visibility.*

(a) *Central Slice* **(b)** *Label Indicated*



the statistics of which provide enough discriminative information to separate the different patterns.

Extracting features from images means finding image transformations according to informational image properties. What information is relevant depends on the task at hand. Methods used for feature extraction from images mainly come from signal detection theory and mathematics (mainly statistical). We recommend Guyon (2006) to get an overview over methods for feature extraction and Lindeberg (2012, 2008, 1994) for a discussion of features in computer vision.

In biomedical image processing, it is difficult to classify organ tissues using only shape or gray level information, because the shape of organs is inconsistent over slices and because intensities overlap considerably for different tissues (Xu and Wunsch, 2005). For these reasons, features often go beyond intensity and include something what can be very generally referred to as texture (see Vyas and Rege, 2006, more precisely a weighted sum of the intensity of surroundings pixels. A survey of features used in Computer-Aided Diagnosis can be found in van Ginneken et al. (2001) abd

Chan et al. (2008).

Methods in image recognition in 3D either take 2D features that work on different viewpoints in the volume, or contain implicit or explicit information about the 3D structure. Examples for the first group of methods are LeCun and Bottou (2004) and Perrotton et al. (2010), for the second group Savarese (2007), and for the third Liebelt et al. (2008) and Liebelt and Schmid (2010). In our work, we took the first approach in using 2D feature maps for classification. This was done recently in a very similar fashion by Danielsson and Carlsson (2011).

Image features used in earlier work on the same dataset by Lopez-Sanchez et al. (2007) were Gabor filters. Previous work, such as reported in Cerquides et al. (2005), and several preliminary tests (not reported) made us expect that a richer feature set tailored for our data set would increase classification performance. The next paragraphs will provide detail on methods that we used for feature extraction. Techniques described in papers I and II and summarized in this chapter are designed to select from this original set of features a (typically smaller) set of features that is most discriminant for a classification task.

It is important to note that features have different properties concerning robustness to transformations and distribution in the volume. On the other side, filters of the scale–space family are shift-, rotation-, and scale–invariant (Florack et al., 1992; Lindeberg, 1994). We tried out Gabor filters, Laplacian pyramids, and luminance and texture contrasts, which come from neuro–psychological research. The idea was that all these features could potentially bring benefit for our task.

Often a too domain-specific adaption of descriptors may lead to less stable representations which are vulnerable to changes in scale and location. Hence, photometric descriptors should be chosen that discriminate well, but that are not too specific to particular regions for parsimony in representation Pantofaru et al. (2006).

3.5.2.1 Laplacian pyramids

The *Laplacian pyramid* (Burt and Adelson, 1983; Crowley and Stern, 1984) is a standard method for compact image encoding where an image is first iteratively smoothed by convolving Gaussians at increasing scales as low-pass filters. This is called the *pyramid of Gaussians*. Then, differences between adjacent scales are

computed.

Formally stated, the two-dimensional Gaussian convolution of the image x, y at scale s is defined as (after Lindeberg, 2012, 2008, 1994)

$$\mathrm{GC}(x; y; s) = \bigtriangledown^2 \left(I(x; y) G_(x; y; s) \right) \tag{3.47}$$

where I is the image, and

$$G(x, y; s) = \frac{1}{2\pi s} e^{-(x^2+y^2)/(2s)} \tag{3.48}$$

is the Gaussian kernel. A level of the Laplacian pyramid is then defined as the difference-of-Gaussians (DOG) (after Burt and Adelson, 1983)

$$\mathrm{LP}(x; y; s; \Delta s) = \mathrm{GC}(x; y; s) - \mathrm{GC}(x; y; s + \Delta s). \tag{3.49}$$

We will refer to these features as Laplacians.

3.5.2.2 Local energy and wavelets

Local energy (Morrone and Owens, 1987; Perona and Malik, 1990) or (equivalently) phase congruency (Kovesi, 1999) in the wavelet or Fourier domain has proved to be more efficient in segmentation than gradient based methods (cf. Kovesi, 2002 for a discussion). A special kind of wavelet are *Gabor filters*, which have received considerable attention because the characteristics of simple cells in the primary visual cortex of some mammals can be approximated by these filters. They are used a lot in pattern recognition and texture segmentation. A Gabor filter is defined as

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}) \cos(2\pi \frac{x'}{\lambda} + \psi), \tag{3.50}$$

where
$$\begin{aligned} x' &= x \cos\theta + y \sin\theta \\ y' &= -x \sin\theta + y \cos\theta. \end{aligned}$$

In equation 3.50, λ represents the wavelength of the cosine factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function in degrees, ψ is the phase offset in degrees, and γ is the spatial aspect ratio, and specifies the

ellipticity of the support of the Gabor function.

Gabor filters, in contrast to the other compared features, incorporate orientation information.

3.5.2.3 Luminance contrast

In eye–tracking studies, Reinagel and Zador (1999) gave evidence for increased *luminance contrast* in fixated regions as compared to control points (fixated points on different images). They defined luminance contrast (LC) as the variance of luminance within a patch (a rectangular patch for practical purposes) divided by the mean intensity of the image. Luminance contrast typically is computed (for most purposes) in patches of 1-2 ° of visual angle (e.g. Acik, 2006).

More formally, given a patch P of pixel intensity from image I and around a pixel (x, y):

$$\mathrm{LC}_P = \frac{\sigma_P}{\mu_P}, \tag{3.51}$$

where μ_P is the mean and σ_P is the standard deviation of intensity of the patch.

3.5.2.4 Texture contrast

By the definition given in Einhäuser et al. (2006), *texture contrast* is the luminance contrast of the luminance contrast of an image. The texture contrast (TC) of a patch is the standard deviation of the luminance contrast values in the patch standardized by the luminance contrast mean of the image. Parkhurst and Niebur (2004) suggested an extension of the saliency map model, predicting a tenfold increased effect of texture contrast compared to luminance contrast. More formally, given a patch $\bar{P}$ from LC_I

$$\mathrm{TC}_P = \frac{\sigma_{LC_P}}{\mu_{LC_P}} = LC_{\bar{P}} \tag{3.52}$$

3.5.2.5 Feature set

The experiments and comparisons that follow in this chapter are based on a set of 177 features, as summarized in table 3.1, and their respective relevance measures and mutual redundancies. The feature set consisted of 127 features from a central image (image index 100). We added 50 probes which have a function in performance

assessment; a good feature selection method should eliminate most of these probes. 49 of these probes were random variables. 25 of those standard normal distributed, 24 uniformly distributed in the interval $(0, 1)$. The last probe consisted entirely of zeros.

In summary, our feature set included (c.f. table 3.1)

- Laplacian maps,
- Gabor features (10 orientations at 10 scales),
- luminance contrast at scales (in pixels) $5, 13, 27, 35, 41, 55, 69, 89, 99$,
- texture contrast (at scales $5, 13, 27, 35, 41, 55, 69$), and
- intensity,
- probes.

Table 3.1: ***Feature set for feature selection.*** *The different features are explained in the text. The last three feature types are so–called probes, useless features introduced for testing purposes. Please, see section 3.5.2 for an explanation of the feature set.*

feature	index range	number
Laplacian Pyramid	$1 - 10$	10
Gabor filters	$11 - 110$	100
Luminance contrast (LC)	$111 - 119$	9
Texture contrast (TC)	$120 - 126$	7
Intensity	127	1
Random std. norm.	$128 - 153$	25
Random uniform.	$154 - 176$	24
Zeros	177	1

3.5.3 Classifiers

We used three classifiers for benchmarking:

- Naïve Bayes,
- GentleBoost, and
- a linear support vector machine (SVM).

As for Naïve Bayes, we relied on our own implementation for multi-valued features using 100 bins[13] for discretization. Given m features $X_1, \dots X_m$ and corresponding targets $Y \in C^n$, classifying a pattern $x = \{x_1, \dots x_m\}$ by Naïve Bayes means

$$\underset{c \in C}{\operatorname{argmax}} \; p(c) \prod_{i=1}^{m} p(x_i | c = Y), \tag{3.53}$$

As for GentleBoost, we used Antonio Torralba's MATLAB toolbox (Russell et al., 2007).

As for SVM (Cortes and Vapnik, 1995), we used libsvm 2.84 (Chang and Lin, 2001), accessed from within MATLAB using an interface by Michael Vogt (Vogt, 2004) from Technical University Darmstadt.

For SVM learning, we tried out several normalization methods. An often used method for SVM (Hsu et al., 2003) is normalizing within $[0, 1]$ or $k[-1, 1]$, which however, has the disadvantage of staunching distributions with outlier values. Graf and Borer (2001) proposed to normalize by the square root of the sum of the squares:

$$\tilde{x} = \frac{X}{\sqrt{\sum X^2}}. \tag{3.54}$$

Comparisons (not shown) showed that classification performance being approximately equal, there was a notable loss of speed with classification after normalization as per Graf and Borer (2001), with z-normalization performing fastest. Consequently, features were z-normalized.

The cost function was made to compensate for unequal class priors, i.e. the weight of the less frequent class was set to $\max\left(\frac{\sharp(Y=c_2)}{\sharp(Y=c_1)}, \frac{\sharp(Y=c_1)}{\sharp(Y=c_2)}\right)$. We set the SVM complexity parameter C to 1 which seemed to be a good choice and on the right order of magnitude.

In preliminary tests (omitted here), we determined that 50 iterations and random sampling seemed to be good choices.

[13] Compare section 3.3.

3.5.4 Experimental conditions

More precisely, the benchmark included the feature selection of the combining redundancy and relevance information using 1. mRmRQ, 2. mRmRD, 3. Greedy, and 4. Hopfield.

Additionally to the methods for selection (c.f. section 3.4), as a baseline, features were selected based on unitary filters, i.e. based on either relevance or redundancy. As for unitary filters, for relevance measures, the s highest relevant features were used and for redundancy measures, at each step the most redundant feature with all the remaining features is removed until the desired numbers of features s are left. Another random baseline selection was computed.

Each relevance and redundancy criterion was evaluated on its own by unitary filters; further all 28 combinations of explained relevance and redundancy measures with the different methods and random selection. Feature sets were evaluated at sizes $[4, 8, 12, 16, 20, 30, 45, 60, 80, 100]$. This choice emphasizes feature sets of sizes ≤ 30 because that was where we found initially the greatest differences between the different methods.

In table 3.2, in the columns, you can find the independent variables of the experiments. The whole set of experimental conditions can be obtained by combining selection schemes with corresponding relevance and redundancy measures, classifiers, and numbers of features. Greedy, Hopfield, mRmRQ, and mRmRD, were tried out with the 28 redundancy and relevance combinations, all classifier, at each number of features. Unitary filters use either a redundancy measures or a relevance measures, but not both, were combined with a classifier and a number of features. The random selection ran with each classifier at each number of features. In total this made up for 3700 experimental conditions.

At each number of features — in order to compromise between sufficient number of validations and acceptable duration of experiments — 10 random samplings of size $n/10$, and for each sampling 5–fold cross-validation was performed. As for random feature selection, 10 random samplings of the data of size $n/10$ were performed and 10 random selections of features in 5-fold cross-validation were tested.

Table 3.2: ***Experimental conditions in the feature selection study***. *Independent experimental variables were selection scheme, redundancy, relevance, classifier, and number of features. lease, compare section 3.4 for an explanation of the feature selection schemes. In total, there are* 3700 *experimental conditions.*

scheme	Red. measure	Rel. measure	Classifier	♯features
Greedy	KSC	SU	Naïve Bayes	4
Hopfield	KSD	CC	GentleBoost	8
mRmRD	RVDM	VDM	SVM	12
mRmRQ	RFC	FC		16
Redundancy filter	CC			20
Relevance filter	JS			30
Random	ST			45
				60
				80
				100

3.5.5 Statistical evaluation

As for the evaluation of the experiments, we relied on Demšar (2006). In that paper, Demsar gives a review of the state of the art of comparing classifiers and provides some recommendations. Accordingly, comparisons of performances of different classifiers should proceed in a one to one fashion, should include performance measures other than accuracy (desirably including ROC or AUC analysis), and use cross-validation. As for tests of statistical relevance, Demsar discusses several statistical tests. He discourages the use of parametric tests in general, such as the paired t-test and argues for the signed rank test for single comparisons. For multiple comparisons of different classifiers he recommends the Friedman test (Friedman, 1937), a non–parametric equivalent of the repeated-measures ANOVA. If the null hypothesis (that performance results are equivalent) is rejected, a post-hoc test, such as the Nemenyi test should test for significant differences between the results.

We used the AUC as our performance measure throughout the analysis as computed by Will Dwinnell's (Dwinnell) MATLAB function. Following the recommendations of Demšar (2006) we did not base the statistics on performances of single

folds but took averages (medians[14]) over folds.

Our statistic tables generally have the following format (by columns):

1. Name of the method; usually the selection scheme followed by redundancy and relevance measures
2. Rank of the method within methods compared in the same table
3. Mean rank of performance (decisive for ordering)
4. Median performance
5. Interquartile range of performance
6. Win-Loss from Friedman test and Nemenyi post-hoc test (F/N)
7. Win-Loss from Wilcoxon signed–rank test (SR)

The first column gives the name of the method, usually the selection scheme followed by redundancy and relevance measures, the second column indicates the rank of the method within methods compared in the same table. Ordering follows by mean rank of performance (third column). Median performance and interquartile range of the vector of performance scores (columns three and four) served for statistical comparisons by the Friedman test and the Nemenyi post-hoc test (F/N), and the Wilcoxon signed–rank test (SR). One-to-one comparisons of methods by these statistical tests can be found in columns five and six as win and loss scores (W/L) indicating statistical significance. For example, in table 3.4 line 3 reads as follows:

	index	**mean rank**	**median**	**iqr**	**F/N W/L**	**SR W/L**
mRmRQ: RVDM+FC	3	18.78	0.98	0.03	69/0	87/0

‘mRmRQ: RVDM+FC“ is minimum redundancy maximum relevance with quo-

[14]According to the central limit theorem, any sum (such as, for example, a performance benchmark), if of finite variance, of many independent identically distributed random features will converge to a Gaussian distribution. This is however, not necessarily to expect for only 5 values, i.e. from 5-folds of cross-validations. After finding partly huge differences between means and medians over cross-validations, in pre-trial runs, we decided to take the more robust median (which in case of normal distributions is equal to the arithmetic mean. As for the error–bar, we plot the *interquartile range* (shortened: *IQR*), which is the difference between values at the first (25%) and the third quartile (75%).

tient measure, RDVM redundancy and FC relevance. Index 3 means it is third ranked method with mean rank 18.78 of 96 methods by AUC. The AUC has its median at 0.98 with an IQR of 0.03. According to Friedman and Nemenyi tests it is better than 69 other compared methods and worse than none. According to the Wilcoxon signed–rank test it is better than 87 methods and worse than none.

Extracts from some tables are given in this chapter, e.g. table 3.4. Tables 3.7 and 3.8 analyze redundancy measures and relevance measures, respectively, over all classifiers, numbers of features, mRmRQ/D, and Hopfield, and relevance or redundancy measures, respectively.

A difficulty in tackling some questions with regard to the Greedy method are the different numbers of features that Greedy produced. Table 3.2 shows the numbers of features used in each combination of redundancy and relevance measure. Because of the difficulties tabulating Greedy with the other methods, we chose not to include this scheme in statistical tests over all numbers of features. However, we included all Greedy schemes in number-of-features specific comparison tables using a threshold of $\frac{|s_{\text{design}} - s_{\text{Greedy}}|}{s_{\text{design}}} \leq 0.1$.

Result tables were computed for each tested number of features all methods, i.e. Greedy, Hopfield network, mRmRQ, mRmRD, unitary redundancy filters, unitary relevance filters with corresponding measures of redundancy and/or relevance. These tables served for overall comparisons of all selection schemes including Greedy, in particular, table 3.3.

The Friedman test, the Nemenyi post–hoc test, and the Wilcoxon signed–rank test are described in the next sections.

3.5.5.1 Friedman test

The Friedman test (sometimes called *two-way analysis on ranks*) is a non–parametric statistical test for consistency of distributions. Given data $\{x_{ij}\}_{n \times k}$ of n blocks and k treatments, each point representing a single observation (the MATLAB function `friedman` allows more), we assign ranks within each row (averaged if tied), obtaining

$\{r_{ij}\}_{n\times k}$, where r_{ij} indicates the rank of x_{ij} within block i. Let

$$r_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} r_{ij} \ , \tag{3.55}$$

$$\bar{r} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} r_{ij} \ , \tag{3.56}$$

$$\mathrm{SS}_t = n\sum_{j=1}^{k}(\bar{r}_{\cdot j} - \bar{r})^2 \tag{3.57}$$

$$\mathrm{SS}_e = \frac{1}{n(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}(r_{ij} - \bar{r})^2 \tag{3.58}$$

Then the test statistic is given by $Q = \frac{\mathrm{SS}_t}{\mathrm{SS}_e}$. For two columns the Friedman test reduces to the Wilcoxon signed–ranks Test. For two rows, it reduces to the Spearman rank correlation coefficient. The p-values can be approximated from a χ^2 distribution. H_0 is that the distributions of ranks within rows are unrelated between columns. If significant values are found, a post-hoc test, such as the Nemenyi test, should be performed (Demšar, 2006).

We used MATLAB's friedman function (MathWorks, 2007b).

3.5.5.2 Nemenyi test

Given data $\{x_{ij}\}_{\times N,k}$, where the number of rows N corresponds to the number of results of each classifier and the number of columns k to the number of classifiers, ranks are assigned within rows, which gives $\{r_{ij}\}_{\times N,k}$. Let R_i be the mean of column i. The statistics of the Nemenyi Test is then for two columns i and j

$$z = \frac{|R_i - R_j|}{\sqrt{\frac{k(k+1)}{6N}}} \ , \tag{3.59}$$

Critical values for these z-values can be found in text books on statistics or in Demšar (2006).

We implemented the Nemenyi test according to his specifications.

3.5.5.3 Wilcoxon signed–ranks test

As an alternative statistical significance criterion we considered the p-values of the *Wilcoxon signed–ranks test* (also called the *Mann-Whitney U test*, henceforth abbreviated *SR*, short for signed rank). This test provides an alternative for the paired t–test when the samples or non–normal. See Ott and Longnecker (p. 319 2008) for a detailed explanation.

We used MATLAB's signrank function (MathWorks, 2007d).

3.6 Results

Statistics were extracted from performance vectors and are given over all three classifiers (Naïve Bayes, GentleBoost, and SVM). Please, note, that there is never "one true way", no single best method ("Rule of Diversity", cf. Raymond, 2003) and we only can give limited recommendations based on the methods that worked well in our experiments. If you want to select features, what is the "best" method depends on how many features you have and how many you want to/have to select, which kind of data you have, and how much time.

The evaluation of results focused on five questions:

1. Which is the best feature selection scheme?
   (a) In particular, are there differences with respect to numbers of features?
2. What are the best measures of relevance and redundancy (RR)?
   (a) What is the best redundancy and relevance (RR) combination?
   (b) What is the best redundancy measure?
   (c) What is the best relevance measure?
3. Do class–conditional distributions give better redundancy estimations?
4. Are the best methods the ones with fewest probes?
5. What is the best feature set?

Question 1 includes feature selection schemes, measures of redundancy and relevance (short: *RR measures*), and combinations of relevance and redundancy. Apart from an overall winner according to our experimental setup, we thought it would be

instructive to find out which selection scheme could give the best results. Important for the evaluation are furthermore differences between the methods with respect to employed redundancy and relevance measures.

As for question 1.a, it was analyzed if relative performance of the different methods is the same when the number of features selected increases. Another issue to be considered is what constitutes a good feature size for the classification task and based on this decision decide on the best selection scheme.

Question 2 concerns comparisons of relevance and redundancy combinations and of redundancy measures (a) and relevance measures (b) among themselves.

Another distinction is whether it helps to take into account class–conditional distributions. As for question 3, it was to be resolved whether this made sense, looking at KSC and KSD redundancy criteria which only differ in using class–conditional distributions and total distributions.

As for question 4, we looked at probe frequency in order to see whether a selection scheme with good performance is automatically one with few probes.

Question 5 deals with the final result of our feature selection. We wanted to find a feature set for classification, so we try to find an answer to this question.

3.6.1 Best selection scheme

Table 3.3: ***Ranking of all the selection schemes****. Please, see section 3.5.5 for an explanation of the statistical evaluation. Abbreviations: F/N – Friedman–Nemenyi, SR – signed–ranks test, W/L – win–loss. Please, refer to section 3.4 for an explanation of the feature selection schemes.*

	index	mean rank	F/N W/L	SR W/L
mRmRD	1	2.10	3/0	4/0
Hopfield	2	2.85	2/0	2/0
Red	3	3.50	1/0	2/0
mRmRQ	4	3.70	1/0	1/1
Rel	5	3.95	1/1	1/1
Greedy	6	5.00	1/2	1/3
rand	7	6.90	0/6	0/6

As mentioned in subsection 3.5.5, Greedy could not be included for some comparisons of totals over all numbers of variables, because of the inconsistent numbers of

features it produced. However, in comparisons with specific numbers of features, we took into account Greedy selections where they produced feature sets of sizes maximally 10% diverging from the sizes of feature sets from other methods. Table 3.3 (table 1 in paper I) was computed from rankings (in contrast to performance) at different numbers of features and includes Greedy. Median performance and iqr were omitted in this table.

According to table 3.3, mRmRD is overall winner followed by Hopfield. The table is ordered by mean ranks (third column). Ranking by medians (fourth column, – in this case – medians over ranks) would have put mRmRQ in second place, however, mRmRQ is by Wilcoxon signed–ranks worse than mRmRD (and Hopfield and unitary redundancy filters are not).

Random feature selection is clearly (and statistically significantly) worse than the all other selection schemes. Greedy is the worst non–random scheme. Unitary redundancy filters come high up at the third place.

	index	mean rank	median	iqr	F/N W/L	SR W/L
Red: RVDM	1	11.73	0.99	0.03	87/0	93/0
mRmRD: CC+VDM	2	18.33	0.99	0.03	71/0	88/0
mRmRQ: RVDM+FC	3	18.78	0.98	0.03	69/0	87/0
mRmRD: RVDM+FC	4	19.92	0.97	0.03	64/1	77/2
mRmRD: RVDM+SU	5	20.35	0.99	0.03	73/1	72/1
mRmRD: JS+FC	6	23.03	0.98	0.03	57/1	67/3
mRmRQ: ST+VDM	7	23.58	0.98	0.03	56/1	78/1
mRmRD: RVDM+VDM	8	24.45	0.97	0.03	51/1	70/2
mRmRQ: KSC+VDM	9	26.72	0.98	0.03	53/2	70/4
mRmRD: JS+VDM	10	26.87	0.97	0.03	52/1	65/4
mRmRQ: RVDM+VDM	11	27.58	0.98	0.03	51/1	69/2
mRmRD: ST+VDM	12	27.78	0.98	0.03	49/3	63/5
mRmRQ: KSD+VDM	13	27.98	0.98	0.03	49/3	63/6
mRmRD: KSD+VDM	14	27.98	0.98	0.03	49/3	63/6
mRmRD: KSC+VDM	15	27.98	0.98	0.03	49/3	63/6
mRmRQ: KSC+FC	16	28.32	0.97	0.03	53/1	55/9
Red: JS	17	28.57	0.98	0.03	49/1	54/2
mRmRQ: KSD+FC	18	28.95	0.97	0.03	55/1	54/10
mRmRQ: ST+FC	19	29.03	0.97	0.03	55/1	53/7
mRmRD: RFC+VDM	20	30.50	0.97	0.03	44/3	51/7
mRmRD: KSC+FC	21	31.50	0.97	0.03	51/3	49/17
mRmRD: KSD+FC	22	31.58	0.97	0.03	52/5	50/17
mRmRQ: JS+VDM	23	31.63	0.98	0.03	43/4	61/3

Continued on next page

Table 3.4 – continued from previous page

	index	mean rank	median	iqr	F/N W/L	SR W/L
mRmRQ: CC+VDM	24	32.23	0.98	0.04	39/3	58/2
mRmRD: ST+FC	25	32.58	0.97	0.03	47/3	45/18
Hopfield: RVDM+VDM	26	34.35	0.98	0.03	42/12	53/14
Hopfield: CC+VDM	27	34.35	0.98	0.03	42/12	53/14
Hopfield: JS+VDM	28	34.35	0.98	0.03	42/12	53/14
Hopfield: RFC+VDM	29	34.35	0.98	0.03	42/12	53/14
...	...	...	...	...	...	...
mRmRQ: JS+SU	80	78.68	0.89	0.09	4/75	15/74
mRmRQ: ST+SU	81	80.75	0.89	0.12	14/79	15/79
mRmRD: ST+SU	82	83.38	0.87	0.11	9/79	13/81
mRmRQ: KSC+SU	83	84.22	0.86	0.12	8/80	12/81
mRmRD: KSC+SU	84	84.77	0.86	0.12	9/80	10/82
mRmRD: RFC+SU	85	85.23	0.82	0.21	3/80	6/84
mRmRQ: KSD+SU	86	85.55	0.86	0.12	8/80	8/84
mRmRD: KSD+SU	87	85.63	0.86	0.12	8/81	8/85
Rel: SU	88	86.27	0.86	0.12	7/81	6/84
Hopfield: RVDM+SU	89	88.17	0.85	0.12	4/86	3/87
Hopfield: CC+SU	90	88.32	0.85	0.12	4/86	3/87
Hopfield: RFC+SU	91	88.32	0.85	0.12	4/86	3/87
mRmRD: CC+SU	92	88.70	0.85	0.17	2/85	2/86
mRmRQ: RVDM+SU	93	89.85	0.85	0.10	3/90	3/86
Red: CC	94	91.37	0.70	0.40	1/92	1/93
mRmRQ: CC+SU	95	93.03	0.79	0.19	1/94	1/93
mRmRQ: RFC+SU	96	94.37	0.68	0.20	0/94	0/94

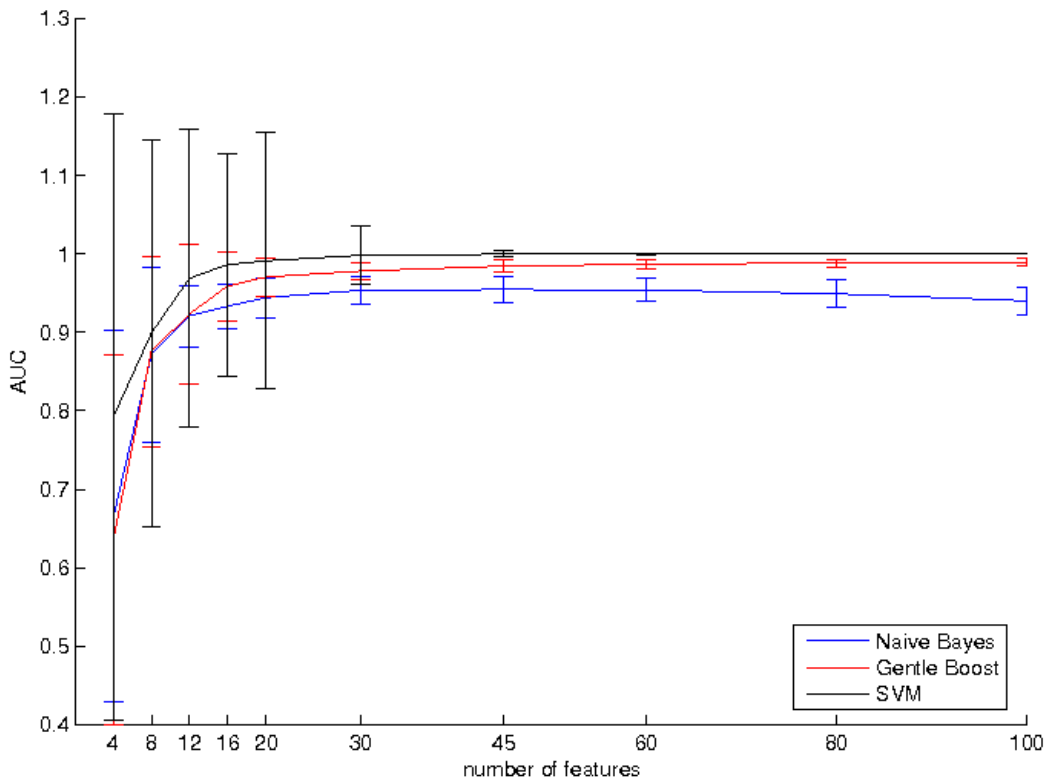
Table 3.4: *Selection schemes over all numbers of features and classifiers. Please, see section 3.5.5 for an explanation about the statistical methodology.*

Table 3.4 shows an analysis over all the classifiers and all the features, excluding Greedy. Unitary RVDM redundancy filter ranks first, followed by mRmRD with CC redundancy and VDM relevance and mRmRQ with RVDM and FC. mRmRQ/D came up on top and mostly better than Hopfield using identical RR measures. In this table, the first Hopfield method ranked 26 (RVDM+VDM). However, many mRmRQ methods are situated in the bottom rows (which explains why on average over all RR measures they showed up worse than the Hopfield selection method in table 3.3).

3.6.1.1 Best selection method by numbers of features

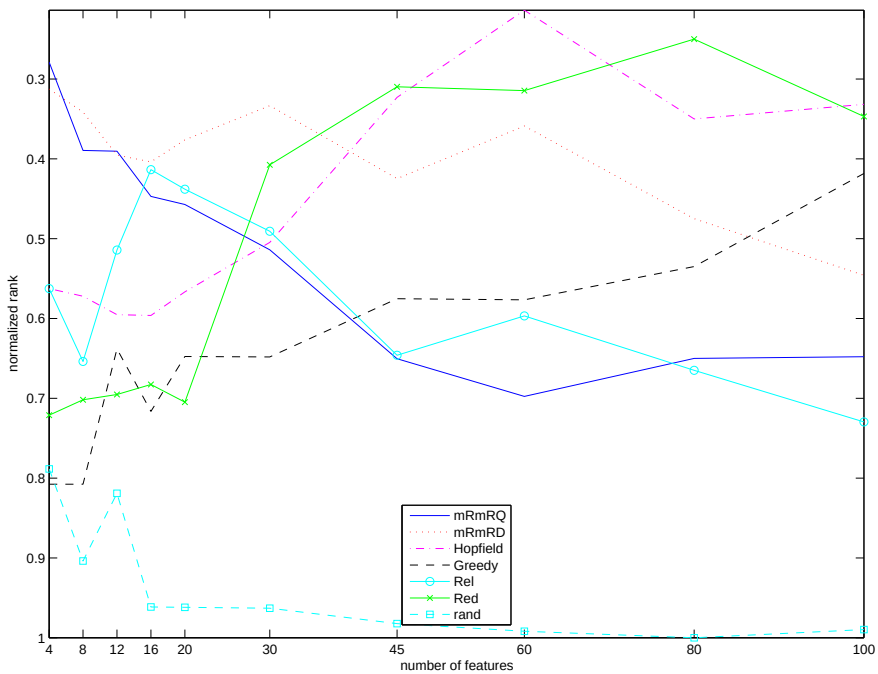
Fig. 3.6 shows the AUC of random selection at different numbers of features. Performance was best with the SVM classifier, second best with GentleBoost, and worst with Naïve Bayes. The clearest differences occurred until about 30 numbers of features. It can be seen however, that the performance followed a curve that was initially log–shaped, but mostly saturated at 100 features. We find the best feature combination and the best combination method at the upper range of the numbers of features.

Figure 3.6: ***Results from random selection.*** *This figure shows the classification performance when features were randomly selected. This is a baseline condition.*



At each number of features, all the selection schemes were ranked. The medians

Figure 3.7: ***Feature selection results: all selection schemes.*** *normalized median rankings of all the selection schemes. The abscissa shows the size of the feature set. A high rank (low value) means that a selection scheme was one of the best in the comparison, a low rank (high value) means that the selection scheme was bad among the compared methods. Some selection schemes are among the best choices for few features, while others outperform for more features. For few features mRmRQ and unitary–filter relevance selection are the best compared. mRmRD, Hopfield network selection, and surprisingly, the selection based only on redundancy are the methods that perform best for many features. Greedy selection improves steadily to become one of the best method for many features. Random selection is the worst compared method at nearly all sizes.*



for each selection scheme are depicted in fig. 3.7. These data served for a global analysis of all the selection schemes across all redundancy and relevance measures which can be found in table 3.3.

Fig. 3.7 sums up these results over all the selection methods. We see that with

Table 3.5: ***Correlations of relevance measures.*** *Abbreviations: FC – fit criterion, SU — symmetric uncertainty, VDM — value difference metric, CC — Spearman rank correlation. Please, see section 3.1 for an explanation of the relevance measures.*

	Spearman correlations			
	FC	SU	VDM	CC
FC	1.00			
SU	0.22	1.00		
VDM	0.68	0.31	1.00	
CC	0.28	0.68	0.41	1.00

more features all the selection schemes become better than the random choice because of the inclusion of less probes. MRmRQ starts as the best method at 4 variables. When adding more features, it improves comparatively less than most other selection schemes. Hopfield, unitary redundancy filters, and Greedy lead to the best improvements as compared to other methods mRmRQ/D and unitary relevance filters.

Generally, Hopfield coped better with higher feature spaces than the other methods and constitutes the best method from 60 features on.

3.6.2 Redundancy and relevance measures

3.6.2.1 Relevance

Table 3.5 shows Spearman correlation coefficients between different relevance scores. It turns out that the relevance measures differ greatly with respect to the importance they assign to different features. VDM and FC, and SU and CC demonstrated large correlations ($\rho > 0.65$).

Fig. 3.8 gives the relevances attributed to each kind of feature by each relevance measure. All the relevance measures seem to concur that Laplacians and some Gabor filters are very useful, however, there are huge difference with respect to Gabor filters, LC, and TC. In particular, it can be observed that CC and SU associate lower relevance to some Gabor filters than to some probes.

Figure 3.8: ***Rankings by features and relevance measures***. *The vertical grid is to illustrate the boundaries between the different types of features. A low value in rank (say a high rank) corresponds to a high relevance. Abbreviations: FC – fit criterion, SU – symmetric uncertainty, VDM – value difference metric, CC – Spearman rank correlation. Gabor features and Laplacians were given a high ranking. CC and SU did poor in distinguishing random features (probes).*

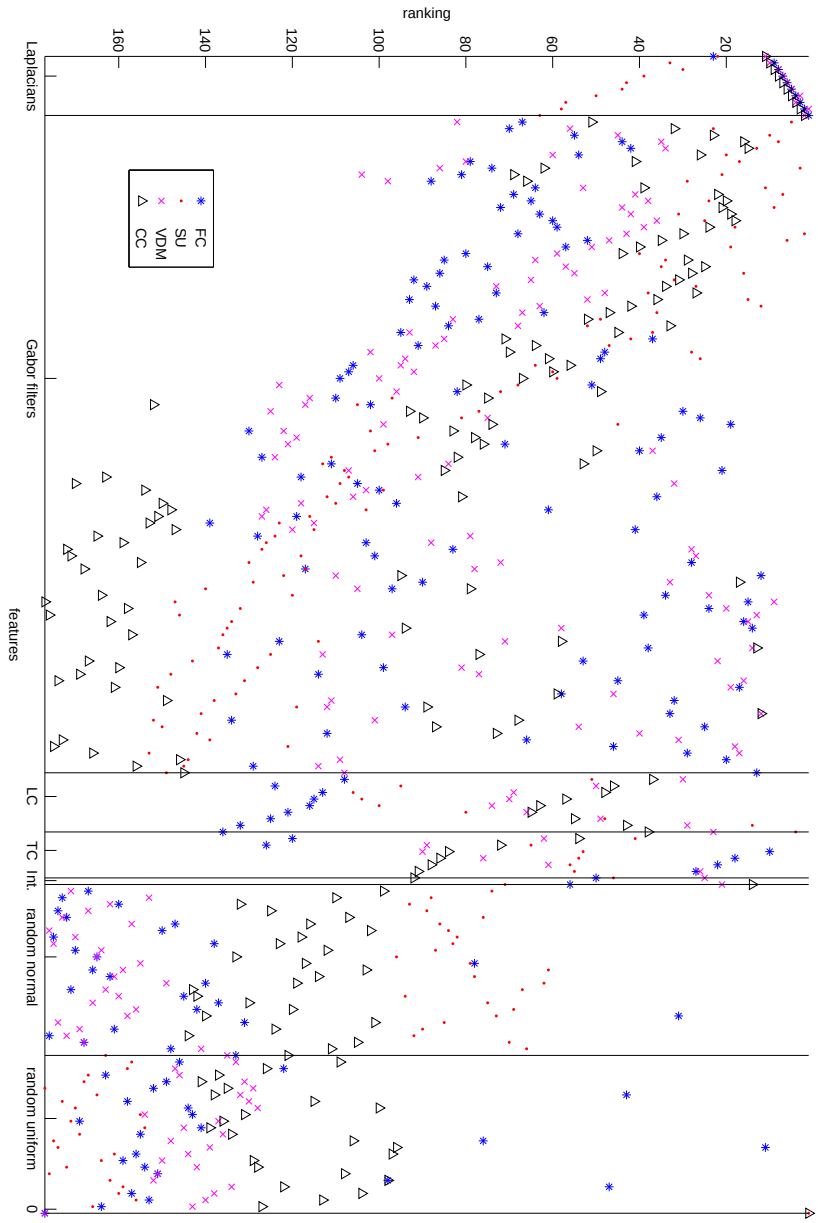


Table 3.6: ***Correlations of redundancy measures.*** *Abbreviations: VDM – value difference metric, KSD – Kolmogorov–Smirnov, CC – Spearman rank correlation, JS – Jensen–Shannon, ST – sign–test, RFC — redundancy fit criterion, KSC — Kolmogorov–Smirnov Class–Conditional. Please, see section 3.2 for an explanation of the redundancy measures.*

Spearman correlations							
	VDM	KSD	CC	JS	ST	RFC	KSC
VDM	1.00						
KSD	0.58	1.00					
CC	−0.19	0.41	1.00				
JS	0.64	0.58	−0.12	1.00			
ST	0.48	0.82	0.22	0.52	1.00		
RFC	0.13	0.42	−0.61	0.09	0.27	1.00	
KSC	0.59	0.96	0.39	0.62	0.85	0.35	1.00

3.6.2.2 Redundancy

Table 3.6 shows the correlations between the redundancy measures. KSC and KSD (unsurprisingly because they are so similar) were found very highly correlating with one-another ($\rho = 0.96$). Both of them also were highly correlated with ST($\rho > 0.8$). CC correlated negatively with some measures, most markedly with RFC ($\rho = -0.61$).

3.6.2.3 Benchmark results

In table 1 of paper II (omitted in this summary), we find a ranking of RR combinations over all numbers of features and over mRmRQ/D and Hopfield. The best combination was RVDM with FC. The table shows nearly coherent groupings by the relevance measure. Everything including SU is clearly at the bottom. Also bad, but better than SU we find combinations with CC relevance. RFC and CC redundancy seem worse than others, with CC having greater deviation. RVDM seems to be a good redundancy measure.

Generally, RVDM and RFC performed very well as unitary filters. Integration of

SU made the performance degrade in many cases with a given redundancy measure when compared to other relevance measures. CC was a bad measure for redundancy; performance was worst when using only CC (Red:CC) and any information helped improve the performance. KSD was also bad, KSC slightly better. Over the different integration schemes, the measures for redundancy and relevance differed in their contribution.

Table 3.7: ***Redundancy measures over all numbers of features and selection schemes.*** *Results exclude greedy (see text). Abbreviations: RVDM – redundancy value difference metric, KSD – Kolmogorov–Smirnov, CC – Spearman rank correlation, JS – Jensen–Shannon, ST – sign–test, RFC – redundancy fit criterion, KSC – Kolmogorov–Smirnov Class–Conditional. Please, see section 3.2 for an explanation of the redundancy measures.*

	index	mean rank	median	iqr	F/N W/L	SR W/L
JS	1	2.68	0.97	0.04	5/0	6/0
RVDM	2	2.94	0.96	0.04	3/0	5/1
ST	3	3.82	0.96	0.07	2/2	3/2
KSC	4	4.23	0.95	0.08	2/2	2/3
KSD	5	4.38	0.95	0.08	1/3	1/4
CC	6	4.54	0.95	0.06	0/2	0/2
RFC	7	5.40	0.95	0.05	0/4	0/5

In table 3.7 we see rankings of the redundancy measures averaged (medians) over mRmRQ/D and Hopfield over all numbers of features. In this case, the clear winner is JS, followed by RVDM and ST together with highly correlated KSC and KSD. RFC comes last, after CC. Both showed only low correlations to the other measures (and highly negatively with each other).

We find a comparison of relevance measures in table 3.8. The statistics is again over mRmRQ/D and Hopfield and over all numbers of features. VDM and FC, which had been found highly correlating, were clearly the best relevance measures. CC comes before SU, which was the clear loser.

3.6.3 Class–Conditional distributions

In section 3.2.1, we introduced two very similar redundancy criteria, KSC and KSD, which both use the Kolmogorov–Smirnov (KS) test. KSD was computed based on

Table 3.8: ***Relevance measures over all numbers of features and selection schemes.*** *Results exclude greedy (see text). Please, see section 3.1 for an explanation of the relevance measures and section 3.5.5 for an explanation of the statistical evaluation. Abbreviations: F/N – Friedman–Nemenyi, SR – signed–ranks test, W/L – win–loss, FC – fit criterion, SU — symmetric uncertainty, VDM – value difference metric, CC – Spearman rank correlation.*

	index	mean rank	median	iqr	F/N W/L	SR W/L
VDM	1	1.49	0.97	0.04	2/0	3/0
FC	2	1.88	0.97	0.06	2/0	2/1
CC	3	2.76	0.95	0.04	1/2	1/2
SU	4	3.86	0.86	0.12	0/3	0/3

the total distributions and KSC on the class–conditional distributions, i.e.

$$KSD(X_1, X_2) = KS(X_1, X_2) \tag{3.60}$$

and

$$KSC(X_1, X_2, Y) = \frac{1}{|C|} \sum_{\forall i \in [1,|C|]} (KS(X_1|Y = c_i, X_2|Y = c_i)), \tag{3.61}$$

where KS refers to the p-values of the KS test as implemented by MATLAB's kstest.

We had introduced both KSC and KSD in order to test, whether it is better to use class–conditional distributions for redundancy estimation. In table 3.6 it can be seen that they were highly correlated (Spearman correlation coefficient of 0.96). Table 3.7 shows the small difference between the two measures could have made a difference in performance with KSC performing better than KSD. The difference in performance is statistically significant according to the Wilcoxon test, but not significant according to the stricter Friedman and Nemenyi tests. We can conclude that estimations based on class–conditional distributions serve equal or better for redundancy measures than estimations based on the distribution totals.

3.6.4 Probes

As explained in section 3.5.2.5, probes are noisy features which we introduced as a quality criterion. Absence of the introduced probes in the selected feature set is a good sign. Fig. 3.9 shows the expected frequency of probes (feature index > 127) in the selection of schemes (over all measures). Frequencies are normalized by numbers of features in the selected set.

We see that Greedy, the worst selection scheme, suffered from probes. Selection based on unitary redundancy filters also had this problem. Unitary relevance filters and Hopfield let in least probes as compared to the other measures. mRmRD/Q were ranked in the middle of the field.

Over all the classifiers, the Spearman correlations of normalized ranks and inverse (subtracting from 1) normalized probe frequencies over all RR combinations at each number of features range between -0.08 and 0.6. This suggests that low probe frequencies are not sufficient for good classifier performance. The correlations follow a curious pattern: they start at medium range (0.5), go down (until -0.08), and climb up again. This suggests low probe frequencies being relatively important at low numbers of variables and when there are few to choose from (but not in-between).

As for probe frequency, in general, the relevance measures, VDM and FC came before CC and SU (which corresponds to their performance ranking). As for redundancy measures, CC lets slip in many probes, which seems to have caused the mediocre performances with CC redundancy. RFC and JS also were more tolerant to probes.

As we will see below in table 3.9, the probe of zeros (feature index 177), enjoyed some popularity. This seems to be a problem that comes from the skewed class-distributions, which makes that 0 can be to 90% associated with one class.

3.6.5 Best features

Table 3.9 lists for all numbers of features the selection method that provided the best performing feature set at that size. From column 3-10 you see the normalized frequencies of features from Laplacian Pyramid (LP), Gabor filters, luminance contrast (LC), texture contrast (TC), intensity (Int.), random normal probes (RN), random uniform probes (RU), and the zero probe. The frequency of each feature type in

Figure 3.9: ***Normalized probe frequencies of all the selection schemes.*** *The abscissa gives the number of features. The scale is inverted so that the selection schemes with least random features are on top. These are relevance and Hopfield network selection. The methods mRmRQ and mrmRD also have little problems with probes. The Greedy and redundancy–based selections have the highest probe frequencies.*

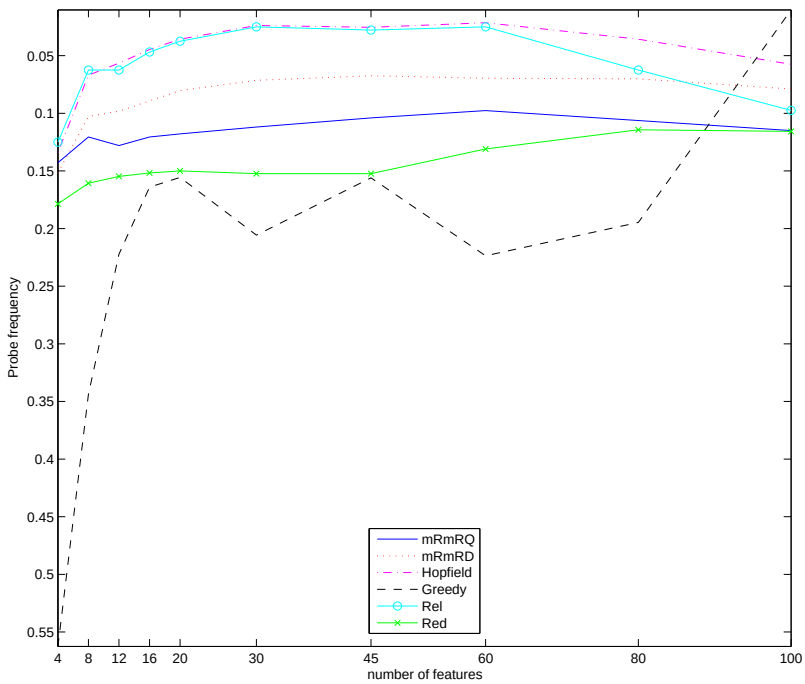


Table 3.9: ***Best selection method for each tested number of features****. This table includes the frequency of each type of feature in the selected set relative to the available number of that type as an indicator for how useful a particular feature type is. Types of features are Laplacians (LP), Gabor filters, luminance contrast (LC), texture contrast (TC), intensity (Int), random normal distributed probes (RN), random uniformly distributed probes (RU). Please, see text or other captions for more abbreviations.*

♯features	selection method	LP	Gabors	LC	TC	Int	RN	RU	Zeros
4	mRmRQ, CC+CC	4.42	0.89	0.00	0.00	0.00	0.00	0.00	44.25
8	mRmRD. CC+VDM	6.64	1.11	0.00	0.00	0.00	0.00	0.00	0.00
12	mRmRD, CC+VDM	7.38	0.89	0.00	2.11	0.00	0.00	0.00	0.00
16	Redundancy Filter VDM	0.00	1.33	4.92	0.00	0.00	0.00	0.00	0.00
20	Redundancy Filter VDM	0.00	1.33	4.92	0.00	0.00	0.00	0.00	0.00
30	mRmRQ RVDM FC	5.31	1.00	1.97	0.84	0.00	0.00	0.00	0.00
45	mRmRQ, RVDM+FC	3.54	1.14	2.62	0.56	0.00	0.00	0.00	0.00
60	Hopfield. KSC+CC	0.00	1.24	2.95	2.95	2.95	0.00	0.00	2.95
80	Hopfield, KSD+FC	0.89	1.28	2.21	2.21	2.21	0.00	0.00	2.21
100	Hopfield, KSD+FC	1.77	1.27	1.77	1.77	1.77	0.00	0.00	1.77

the selected set was divided by the frequency expected from prior probabilities. As for Gabor filters, the a-priori probability is $100/177 \approx 0.56$. For 100 features, the expected number of Gabor features is $0.56 \times 100 = 56$. The figures corresponding to each number of features and feature type tell how much the frequencies found for the type exceed or fall behind expectations. For example, for 100 features from Gabor filters there were 1.27 times more than expected.

Laplacian Pyramids, intensity, texture contrast, and luminance contrast appear prominently (relative to their proportion in the feature set). There were many Gabor filters present. It is remarkable that only few probes are selected (the zeros each time, i.e. feature 177).

3.7 Conclusions

In this chapter, we summarized a framework for measuring the redundancy and relevance of features, which was previously presented in papers I and II, and explained several measures of redundancy and relevance within this framework.

We explained two measures for the evaluation of feature relevance and redundancy that were newly introduced in paper II, *fit criterion* (FC) and *value difference*

metric (VDM). We also presented an application of Hopfield networks to feature selection within the framework of maximum relevance and minimum redundancy.

We explained experiments which were conducted in order to compare the validity of different relevance and redundacy measures and feature selection schemes. The testing involved an exhaustive comparison of all combinations at different sizes of feature sets. Additionally, it included several baselines such as useless features that were introduced to test selection quality (probes), particularly random features, a zero–valued feature. Baselines for selection schemes were random selection and unitary–filter selection.

Of the compared relevance measures, the measures proposed in paper II, the best were the value difference metric (VDM) and the fit criterion (FC). Rank cross–correlations (Spearman correlations) favored the zero–feature (cf. fig. 3.8). Because of its formulation, Spearman rank correlation coefficients are unsuitable for comparisons between distributions with highly different cardinalities, such as the case for comparing classes (cardinality 2) and features (possibly $n \approx 200,000$). Possibly other more appropriate correlation measures should be used instead. As for the redundancy measures, the Jensen–Shannon divergence (JS), the redundancy value difference metric (RVDM), and the sign–test (ST) resulted in good predictions. The redundancy fit criterion (RFC), which is based on the FC relevance measure, may have been too simplistic.

On the whole, for all the tested features we observed that mRmRD was the best combination scheme. Selection with Hopfield networks showed huge improvements for higher–dimensioned feature sets. The redundancy unitary–filter based value difference metric also performed well.

In general, we observed a log–shaped performance curve with feature size. Therefore, we chose the most features (100) as best feature size and the corresponding method as the best method, the Hopfield network with the Kolmogorov–Smirnov Test (KSD) as redundancy measure, and FC relevance (FC).

It would be interesting to compare methods for feature reduction as opposed to feature selection to the results obtained.

Although this study was limited to binary classification, its methods are not. It remains to be seen how our feature selection scales up from a two-class problem to

multi-class domains with thousands of features.

In our experimental design we emphasized few numbers of features (70% of feature sizes below 50), which turned out favorably for mRmRD. The Hopfield network selection seems to perform well for high–dimensional feature spaces and could be used in the analysis of complex data. It remains to test the methods in higher dimensional feature spaces.

The general aim of the presented studies was to extend a classification pipeline for biomedical images by a feature selection in order to make the performance more robust and to allow for a cleaner separation between the classes. Experiments on other data set would have been valuable for validation, however, more extensive testing remains for the future.

It could be that other methods for density estimation could bring an improvement for some measures. Dougherty et al. (1995) tested several methods of discretization and subsequent classification using Naïve Bayes and the C4.5 decision tree. They concluded that an optimization according to the minimum description length principle performed better than other tested methods. More recently, Liu et al. (2002) systematized and tested several methods of discretization and again found the minimum description length performing best. Peng et al. (2005) chose to discretize data using $\mu \pm \alpha\sigma$, with $\alpha \in [0, 2, 0.5]$.

As for relevance and redundancy measures, it is not at all obvious that a single universally best measure exists for all applications (Weeds and Weir, 2003). While this study showed the usefulness of certain metrics and schemes, results should not be taken as universal for all feature selection problems.

Chapter 4

A genetic clustering algorithm

In this chapter, a clustering technique based on a genetic algorithm will be explained, which was introduced in paper III. In the proposed algorithm, the searching capability of genetic algorithms was exploited to find appropriate cluster partitions in the feature space such that a fitness metric of the resulting clusters is optimized. The advantage of this technique was the generality of the admitted fitness functions. For example, we could take distance functions that do not rely on the concept of a cluster centroid. This was meant to allow the clustering to find more complex clusters, as opposed to Gaussian–shaped or hyper–spheroidal ones, which are preferred by more conventional algorithms (see chapter 2). The algorithm was tested on several commonly used machine learning (UCI) datasets, and compared to a standard clustering algorithm (fuzzy c–means). Furthermore, several cluster validity criteria were compared.

The main documentation and results can be found in paper III. In this chapter, the methods and results are briefly summarized.

4.1 Methods

The genetic algorithm has the following characteristics:

- as fitness function, any internal cluster validity measure can be used,

- it incorporates a measure of intra–cluster distance in the cross–over operation (as opposed to just global fitness),
- it uses a mutation bias to find appropriate genes (points) for mutation (comparable to temparature in the *simulated annealing* technique Kirkpatrick et al., 1983),
- at each iteration, it self–adaptively adjusts the global mutation rate in order to keep variance of fitness within a certain range,
- it keeps a small ($M = 20$) population of candidate solutions and employs elitist recombination.

Figure 4.1 shows a pseudo–code that should explain the algorithm.

Given data D and the desired number of clusters k, the algorithm returns a cluster assignment for each point in a membership matrix $U \in \{1, k\}^n$. The population consists of K individual candidate solutions, which are crisp cluster partitions in terms of membership matrices $U^K = \{U_1^n, \ldots, U_K^n\}$, where n is the number of points in our dataset, $U_i^j \in [1, k]$ and K is the number of individuals or chromosomes for each generation. Before the first iteration, one random parent is initialized and K children are created from random permutations (`mutate()`).

For speed–up, the algorithm uses several heuristics. These include the following:

- the inclusion of a mutation bias,
- the preferred selection of individuals with desired traits (good assignment of a cluster), and
- adjustment of the learning rate according to the history of the fitness.

For permutations, a matrix M was taken as a mutation bias for each point. M regulates the mutation probability for each point in the dataset. At each iteration this matrix $M \in [0, 1]^{n,k}$ is updated according to the success and failure of cluster assignments in the previous generation, giving each point its share proportional to the size of the cluster. In the beginning, mutations for any point to any cluster assignment are assumed to be equally probable. Over time, this matrix should converge so that the arguments of the maximum for each point reflect the cluster assignment. In this way, the adaptation of M is a simulated annealing. The mutation

Figure 4.1: *The genetic algorithm with biased mutation and local cross–over (simplified description). Input parameters are data D and number of clusters k. Parameters within* learningrateupdata() *are* $\alpha_{\min} \leftarrow 0.01$*,* $\alpha_{\max} \leftarrow 0.5$ *and* $\alpha_{\text{inc}} \leftarrow 0.05$*. The stop criterion is explained in the text.*

Input: D, k
Output: U, M
`// Parameter definitions:`
maxiteration $\leftarrow 5000$, $K \leftarrow 20$, minchange $\leftarrow 0.0001$,
$\alpha \leftarrow 0.1$, $i \leftarrow 0$,
`// The mutation matrix of size` $n \times k$ `is set to equal values and normalized to row sum 1:`
$M \leftarrow 1^{n,k} \times 1/k$
`// Algorithm starts:`
$U \leftarrow \text{newgeneration}(M, K, \alpha)$
while $i <$ maxiteration *and not* stop **do**
  $(F, intra)^K \leftarrow \text{fitness}(D, U)$
  if $i > 0$ **then**
    $\text{learningrateupdate}(F, F_{\text{old}}, \alpha)$
  else
  end
  `// Select the fittest individual:`
  $w_1 \leftarrow \text{argmax}_i F^i$
  `// Select the individual with the best single cluster:`
  $(w_2, \text{wc}) \leftarrow \text{argmax}_{i,j} \text{intra}^{i,j}$
  `// Select the individual that is best for a random cluster:`
  $\text{rc} \leftarrow \text{ceil}(\text{rand} \times k)$
  $w_3 \leftarrow \text{argmax}_i \text{intra}^{i,\text{rc}}$
  $U1 \leftarrow U^{w_1}$
  $U2 \leftarrow \text{crossover}(U1, U^{w_2}, \text{wc})$
  $U3 \leftarrow \text{crossover}(U1, U^{w_3}, \text{rc})$
  $M \leftarrow \text{mutationbiasupdate}(F, intra, M)$
  $U \leftarrow$
  $\text{mutate}(U1, K/2, M, \alpha) \cup \text{mutate}(U2, K/4, M, \alpha) \cup \text{mutate}(U3, K/4, M, \alpha)$
  $i \leftarrow i + 1$
end

bias, or *temperature*, of each point is indicative of the algorithms confidence in its assignment, similar to a fuzzy cluster membership matrix.

At each iteration, the winner `U1` was determined as the individual from U^K that obtained the highest fitness according to a given cluster validity function. This individual was then permuted to yield $K/2$ individuals. For the rest of $K/2$ individuals in the next generation, we used the following procedure: for each cluster assignment, we evaluated the individual with the lowest intra–cluster distance (or inversely, the highest intra–cluster similarity). We took the individual with the best value for intra–cluster distance, cross-combined it with the overall winner, and permuted the result to yield $K/4$ additional individuals. We then randomly took one of the individuals with low intra–cluster distance to cross-combine with the overall winner and again permuted to obtain $K/4$ individuals. Thereby, each new generation consists of K individuals.

The learning rate was increased when the fitness function could not distinguish between candidates and decreased when the previous winner was best again. Learning or mutation rate α was regulated within a fixed range, between $\alpha_{\min}$ and $\alpha_{\max}$. If the previous winner came up a second time as winner, this could mean there were too many mutations and α was down-regulated. If, on the other hand, changes were so minuscule that the chromosomes could not be distinguished by the fitness function (in our implementation this could be the case when the winner is 1), α is up-regulated. We took the same value, α_{inc} for linear increments and decrements.

We concluded in pre–trials (not reported in this thesis) that the biased mutation operator gives a significant speed up for optimization. We used as stopping criterion (`stop`), a threshold `minchange` which is compared to changes to the mutation bias operator M. If the changes per point of M did not reached twice in a row this threshold, the algorithm was terminated. Otherwise the algorithm iterated until `maxiteration`.

As fitness function we were at liberty to choose many different cluster validation functions. In section 2.1.1 of this thesis, a brief overview over such measures was presented. In the following paragraphs the fitness functions that we tried in paper III are presented.

4.1.1 Euclidean distance

The fuzzy c-means and k-means algorithms use the Euclidean distance from centroids as criterion function. The distance of a point x to centroid i, $c_i = \{c_{i,1}, \ldots, c_{i,m}\}$ (m *features* or *dimensions*) is defined as follows:

$$d_{\text{Euclidean}}(c_i, x) = \sqrt{\sum_{j}^{m}(c_{i,j} - x_j)^2} \tag{4.1}$$

where c_i, the centroid, is adapted at each iteration to the center of gravity of cluster i. Points should be assigned to a cluster such that distances to its cluster centroids are small, while distances to centroids of other clusters are large.

4.1.2 Global and local Mahalanobis distances

The Mahalanobis distance (Mahalanobis, 1936) is a distance measure capable of dealing with hyper–elipsoidal distances, as opposed to hyper–spherical distances (e.g. Euclidean). The point–distance from cluster centroids would be defined as:

$$d_{\text{Mahalanobis}}(c_i, x) = \sqrt{(x - c_i)^T S^{-1}(x - c)} \tag{4.2}$$

where S is the covariance matrix usually corresponds to D, however there have been also experiments with taking cluster covariances. It has been shown experimentally that taking local and global results can make a big difference (Lebart, 1992).

We implemented the Mahalanobis distance for our algorithms in two variants, with global and local (cluster–specific) covariances, which we will refer to as *global* and *local Mahalanobis*.

The computation of the inverse of the covariance matrix presented a problem with few data points, because it easily becomes singular or near singular. This was especially true for the local version, where the covariance is computed from points of each cluster. The inverse of the covariance matrix could take extreme values. To prevent this from happening, a restriction was implemented, so that clusters had to have at least the size $\frac{(k-1)}{k} * \frac{n}{k}$, where k is the number of clusters and n the number of points in the dataset. Smaller clusters were heavily penalized. This

restriction was implemented for all used cluster validity measures in order to have them comparable.

4.1.3 SVD entropy

Alter et al. (2000) proposed an entropy measure based on the distribution of eigenvalues. This entropy has found application in different areas including feature filtering (Varshavsky et al., 2006). However, we are not aware of any previous application to clustering. The main idea of the entropy criterion is to find clusters with points that have a low entropy and, at the same time, clusters that have a high entropy when joined with other clusters.

Following the definition by Varshavsky et al. (2006), formally, if s_j denotes the singular values of the matrix A, s_2 are the eigenvalues of the $n \times n$ matrix AA^t. The normalized relative values are given as:

$$V_j = \frac{s_j^2}{\sum_k s_k^2} \tag{4.3}$$

and the dataset entropy as:

$$H = -\frac{1}{\log n} \sum_{j=1}^{n} V_j \log V_j \tag{4.4}$$

This entropy takes values in the range $[0, 1]$. $H = 0$ stand for an ultra-ordered dataset that can be explained by a single eigenvector and $H = 1$ corresponds to a disordered matrix with a uniformly distributed spectrum. We used this as our intra–cluster distance.

As a measure of distance between two clusters, the inter–cluster distance, we used this SVD entropy:

$$H_d = H_{\text{all}} - (H_1 + H_2) \tag{4.5}$$

where H_1 and H_2 correspond to the SVD entropies of clusters C_1 and C_2 and H_{all} to the entropy of the combined cluster $C_1 \cup C_2$.

As a fitness function we combined linearly the thus obtained intra– and inter–cluster validities.

4.2 Results

We applied our algorithm to several data sets from the UCI machine repository (Frank and Asuncion, 2010). These were the Wine, Iris, Ionosphere, and Breast data sets. We randomly initialized the candidates, set K to 20, and set the maximum iterations for our algorithm to 5000. The clustering of all data sets was completed within several minutes on an off-the-shelf desktop office computer.

We used the Jaccard index (c.f. equation 2.4) to quantify the correctness of results from the genetic clustering algorithm by comparing the correct cluster assignment and the obtained clusters from the algorithm. A score of 1 would mean that all points were correctly assigned.

We compared our results to results from the fuzzy c-means algorithm, which next to k-means is probably the most popular algorithms for clustering. Fuzzy c-means (Bezdek et al., 1984) is an improved version of k–means, which is more robust to outliers and overlap (e.g. Mingoti and Lima, 2006).

In paper III it is shown how starting from random initialization the genetic algorithm optimized according to different fitness functions. Table 4.1 compares clustering performance by fuzzy c–means clustering (FCM) and the genetic algorithm (GA) based on four different fitness functions, Euclidean distance (Eucl), global Mahalanobis (MaG), local Mahalanobis (MaL), and SVD Entropy (Ent). The values displayed correspond to averages (medians) over 10 runs. The genetic algorithm was run for 5000 iterations. The values stand for the coincidence of found clusters (Jaccard index) with the real cluster assignments, so that high values indicate good solutions (1 for the perfect solution).

For the Iris dataset, we obtained a high coincidence with the true cluster assignments with the entropy fitness criterion. The fuzzy c-means algorithms in its default implementation in MATLAB, with fuzzifier parameter $m = 2$, achieved a Jaccard score of 0.6959.

4.3 Discussion

In paper III, a genetic algorithm for clustering was proposed and tested. Optimization criteria consisted of information theoretic measures on complete clusters (SVD

Table 4.1: ***Comparison of clustering Results****: Clustering results of the fuzzy c–means algorithm and the genetic algorithm based on four fitness functions. Units express the coincidences between correct partitions and found partitions (Jaccard Index). FCM stands for fuzzy c–means, GA for the genetic algorithm. As for the fitness functions employed for optimization in the genetic algorithm, Eucl stands as Euclidean distance, MaG for the Mahalanobis distance with global covariance, MaL for the Mahalanobis distance with local (i.e. cluster) covariance, and Ent for SVD entropy. The best results for each dataset are typeset in bold.*

Data	FCM	GA-Eucl	GA-MaG	GA-MaL	GA-Ent
Breast	0.5219	0.4035	0.6564	**0.6750**	0.4008
Wine	0.4120	0.4160	0.4261	**0.5741**	0.3565
Ionosphere	0.4314	0.4176	0.4544	**0.5606**	0.4040
Iris	0.6959	**0.6997**	**0.6997**	0.5779	0.4677

entropy), the Mahalanobis distance with global and local covariance, respectively, and the Euclidean distance. Results for real-life datasets were presented and compared to partitions obtained by fuzzy c–means. The dataset comprised different data distributions.

We think that the results show that some of the problems are very difficult, where popular off–the–shelf algorithms such as the fuzzy c-means do not achieve good results. We conjecture that the unsatisfactory results for fuzzy c–means reflect in part the fact that Euclidean distances from the centroids are not very sensitive and might even be inappropriate for some datasets. We think that our results underline the importance of finding a fitness function that is both sensitive to small changes in partitions and works for the particular dataset (c.f. Jain and Dubes, 1988).

It was shown that — given a fitness function which is appropriate for the data — the algorithm can converge to good solutions. We conclude that the results from clustering using our genetic algorithm were competitive with results from fuzzy c–means. Our good results reflect in part the use of proper fitness functions, the application of which was made possible by our genetic algorithm. We tested the algorithm with Euclidean distances, Mahalanobis distances, but we could have used other validity measures, as well. Results with the SVD Entropy measure were generally disappointing, performing worse than the fuzzy c–means algorithm in our

tests. We think that both Mahalanobis distances (local and global versions) gave very promising results. The local version returned at least satisfactory results for all data sets, achieving the top result for Breast, while the global version performed well for Iris and Breast.

In practice, the algorithm often converged fast. We think that the self–adaptation of the mutation rate helped greatly to find solutions. We conjecture that this is because it helps to maintain a certain level of correlation of fitness between candidates (c.f. Altenberg, 1995). We think that another reason for this good convergence is the cross–over operation, where we applied a goodness value for each cluster. We are not aware of a previous application of intra–cluster validity to the crossover operation in genetic algorithms.

Preliminary experiments have been performed where this genetic algorithm was extended to run without fixing the parameter k in advance. The inter–cluster distance was used to merge or split clusters. Further, by extensions of the mutation routine new clusters could appear or old clusters could disappear.

We want to emphasize that results from our algorithms were achieved with a global set of parameters. We assume that parameter optimization could yield better results. The algorithm did not converge on a good solution for some combinations of data sets and fitness functions. This goes to show that there is still work to do and therefore we see these results as preliminary. There is also still testing to be done on more datasets, however, we think that results presented in paper III are promising.

We think it can be an interesting possibility to initialize the cluster partitions by a conventional clustering algorithm and then run a genetic algorithm for refinement.

Part II

Representations of odors in the early olfactory pathway

Chapter 5

The sense of smell

> Zitto! mi pare sentir odor di femmina! — Don Giovanni in Mozart's eponymous play.

Olfaction, more commonly called *the sense of smell*, gives us information about the presence of small airborne substances in our proximal and distant surroundings. Together with gustation, the sense of taste, it is one of the two chemical senses in humans and most vertebrates. This chapter deals with the importance of the sense of smell and with its biology in vertebrates – especially rodents, particularly rats. The main interest therein concentrates on the olfactory bulb and its glomerular layer. We will briefly discuss the relationship of molecular structure to smell and how sensory impressions of smells are stored and processed in the brain. Chapter 6 will then deal with representations in the olfactory bulb.

5.1 The importance of smell

Olfaction is an evolutionary very primitive sense (c.f. Dryer, 2000; Eisthen, 1997). Historically, the sense of smell has been considered as a second class sense, especially in humans, as compared to other senses such as vision or hearing (cf. Shepherd, 2004; Le Guérer, 2002; Howes, 1991). In fact, when analyzing data across many species, Barton et al. (1995) found that volumes of visual and olfactory areas are negatively correlated and they suggested that this negative correlation indicates a trade–off

between vision and olfaction. The argument is, however, more complicated than vision supplanting olfaction (c.f. Smith et al., 2007), particularly since comparisons are often made relative in terms of volumes relative to the total brain, which is problematic, since it is unclear if different senses exhibit similar relationships to body size (Smith and Bhatnagar, 2004).

The comparatively low number of functional receptor genes and the high proportion of pseudogenes in humans compared to other vertebrates (Aloni et al., 2006) have been taken to support the case that humans perform worse in different tasks related to smelling. It is a general assumption that size of brain structure is related to function (e.g. Kaas, 2000) and the analogous assumption with the number of functional receptors seems parsimonious. Regarding sizes of regions this has been experimentally validated in studies on humans regarding the size of the olfactory bulb and the olfactory sulcus (Buschhüter et al., 2008; Hummel et al., 2003).

Surprisingly, Maresh et al. (2008) found more glomeruli in humans than had been previously found in other animals, which could then predict actually a superior functionality of the human OB. Laska et al. (2005) investigated in different species olfactory discrimination with enantiomers, i.e. identical molecules with varying structure, and they found little correlation between the number of active receptor genes and performance, however, it is not clear if this result generalizes to a broader set of odorants (c.f. 5.3.1; Keller and Vosshall, 2008).

Obviously, environmental constraints, such as diet and ecological niche play an important role in determining how much a species has to rely on smell (Barton et al., 1995). Asahina et al. (2008) argued that olfaction gives fruit flies a survival advantage and the argument could be extended for other animals, as we will discuss. In fact, more and more studies reveal the importance of olfaction in many different animals, including humans.

The sense of smell is important for many aspects of life. These can be grouped into four, where the first three certainly apply to humans: (c.f. Gottfried, 2006; Wilson and Stevenson, 2006, chapter 5)

- food — detection of food sources and control of food quality,
- social communication — recognition of kin, identifying mates,
- threat — avoidance of dangerous conditions (including predators), and

- navigation — homing and territorial marking.

Therefore, the sense of smell is crucial for survival in many animals, including many for behaviors such as maternal bonding, mating, kinship recognition, territorial defense, and aggressive behavior (c.f. Sanchez-Andrade and Kendrick, 2009; Tirindelli et al., 2009; Savic et al., 2001; Martin et al., 2011; Schleich and Zenuto, 2010; Dehnhard, 2011). While the mechanisms of information processing that underlie these behaviors are still poorly understood, they seem remarkably similar across phyla and species (Kaupp, 2010; Zhang and Firestein, 2009; Ache and Young, 2005; Strausfeld and Hildebrand, 1999; Hildebrand and Shepherd, 1997). Some of these similarities could come from common evolutionary ancestry, others might reflect functional convergence that resulted from pressure by similar constraints and tasks (Eisthen, 2002). This indicates that components of the olfactory system, by which odorant molecules are sensed by peripheral receptors and then translated into neural activation and into odor percepts, have remained consistent over millions of years and throughout ecological niches.

Smell is also important in humans, as we will discuss now. While human subjects report not being significantly affected in their lives after olfactory loss; we are often not aware of the effects that smell has on our lives. Patients with a loss of smell, after a head injury or an infection, do not only lose a great part of their flavor perception, but also tend to suffer more from depression and are affected in their psychological well–being and sex lives (Hummel and Nordin, 2005). This could indicate that olfaction often works on an unconscious level. Consistent with this interpretation is that we recognize smells but often have problems labeling them linguistically, which is sometimes called the *tip–of–the–nose phenomenon* (Cleary et al., 2010).

Olfactory memories are associated with emotional arousal (c.f. Willander and Larsson, 2007) and as we will discuss in below, in section 5.4.2, the salient affective dimension is characteristic for smell. This could have to do with the fact that olfactory stimuli are transmitted to the amygdala (Zald and Pardo, 1997), among other structures, which is associated with emotional processing (Ledoux, 2008) and evaluation of biological significance (Pessoa and Adolphs, 2010). It is anecdotal wisdom, voiced in Marcel Proust's *À la recherche du temps perdu* that smell can evoke vivid memories from early life. Indeed, Maria Larsson and colleagues found that,

contrary to visual and verbal cues, olfactory cues tend to bring back memories from early in life (Larsson and Willander, 2009; Willander and Larsson, 2006). While it is possible that olfactory memory follows a slower decay function, another explanation for the findings of age distribution of memories could be the complex nature of olfactory experience which implies that we rarely have the same sensory impressions repeated and therefore olfactory memories and therefore they could be less prone to be erased or overwritten (c.f. Lewandowsky et al., 2009).

Generally, it is assumed that there exist two classes of receptors, 1. olfactory receptors (ORs) and 2. pheromone receptors. These two types are situated in their respective anatomically separated areas in the nose, i.e. main olfactory epithelium and vomeronasal organ (also called *Jacobsen's organ*). This crisp distinction, however, is not without challenge (Sam et al., 2001; Spehr, 2010). Stowers and Logan (2010) proposed a distinction between three types of olfactory stimuli, each of which with their specific sites of detection in the nose and respective sites of projection in the brain:

- pheromones — which trigger emotional and behavioral responses in conspecifics,
- instinctive cues — which elicit conserved, stereotyped behavior on first and subsequent exposure (such as spoiled food),
- odorants — chemicals that have a perceptual quality and which can be associated with behavior given a learning process.

Pheromones (from Greek phero "to bear" and hormone "urge") are substances which are secreted by an animal and have an effect on the behavior or the physiology of conspecifics (Karlson and Luscher, 1959). For an in–depth treatment of pheromones, refer to Wyatt (2003); Dulac and Torello (2003) and for a recent review to Wyatt (2010).

In many species, pheromonal compounds have been identified (e.g. Schaal et al., 2003). Behavioral impacts of odorants in humans are known, some of which could be pheromonal, according to the definition above. In contrast, although there are candidates, such as, for example, androstenone (Dorries et al., 1989; Hays, 2003), the identification of human pheromonal compounds is a matter of debate and, according to Wyatt (2009), none has been conclusively identified yet.

To give examples for behavioral effects of odorants in humans, it is known that odor signals mediate synchronization of menstrual cycles among women (Stern and McClintock, 1998) and that newborn babies can find their mother's nipples by smell (Varendi et al., 1994), which could be regarded as olfactory imprinting in humans. Odors also play a role in mate selection (Moshkin et al., 2011), social preferences of faces (Li et al., 2007; Todrank et al., 1995), and can help us to recognize each other by kinship (Porter et al., 1986; Hold and Schleidt, 1977). Odors are important for emotions, for example, certain kinds of odors attenuate stress responses (Nakashima et al., 2004). Also aggression, fear, and stress responses can be triggered by smell (Mujica-Parodi et al., 2009; Albrecht et al., 2011; Haegler et al., 2010). Tears reduce sexual arousal in men by a chemical signal (Gelstein et al., 2011), and putative sex hormones can trigger hypothalamic responses associated with sexual arousal in humans of the opposite gender (Savic and Berglund, 2010; Kovács et al., 2004; Savic et al., 2001).

The main pheromone processing system, the vomeronasal system, is likely not functional in humans (Trotier, 2011), a fact which leaves some mystery about the signaling process involved in putative pheromone effects, some of which were mentioned above. It is known, however, that at least some pheromones in vertebrates could be detected via the main olfactory epithelium (Wang et al., 2007; Liberles and Buck, 2006; Rodriguez et al., 2000).

Olfaction is also important for food quality evaluation (e.g. Warwick et al., 1993; Lawless, 1991; Massler, 1980). Odors influence our perception of taste over the retronasal pathway (Valentin et al., 2006; Mozell et al., 1969; Finck, 1913)[1] and if we taste food without smelling it (when nasal passages are blocked), we have significantly more difficulty in identifying flavors (Mozell et al., 1969). This means that sensory qualities of foods are partly determined by smell. Conveniently, the English language provides the word *flavor*, as combined perception from taste, olfactory, and additionally the trigeminal system (Berridge and Fentress, 1985). There is not only direct olfactory evaluation, but Galef and Giraldeau (2001) additionally found in rats that smelling the breath of a poisoned animal can induce avoidance behavior

[1] And taste influences odor perception (Green et al., 2012).

of the ingested foods.

Apart from social communication as well as locating and identifying food, the sense of smell is useful in warning us of danger, for example to detect fires (Lloyd and Roen, 2002; Chalke et al., 1958) or toxic substances (Ruth, 1986).

In clinical research, olfactory capabilities (or loss thereof) have been associated with neurologic and psychiatric diseases. High detection thresholds for smells, together with impaired identification, can be an early indicator for the onset of Alzheimer's disease and cognitive decline (Doty, 2009; Olofsson et al., 2009; Djordjevic et al., 2008).

In order to perform functions underlying aforementioned behaviors, animals have to create meaningful perceptual representations of odor qualities and this will be dealt with in chapter 6.

5.2 Commercial applications

As discussed in the review by Gilbert and Firestein (2002), there are many commercial applications in olfaction.

One important application are gas sensors which detect the presence of gases within an area, often within applications for safety. They are mostly based on metal oxide (MOX) or other materials, such as conductive polymers. The term *machine olfaction* describes automated systems that measure the concentrations of airborne molecules using sensor technology and pattern recognition algorithms (c.f. Gutierrez-Osuna, 2002a). With reference to biology, these systems are sometimes called *artificial noses* or *electronic noses* (c.f. Röck et al., 2008; Hierlemann and Gutierrez-Osuna, 2008; Persaud and Dodd, 1982). On the sensor and software level, they are different to biological systems in several aspects (Berna et al., 2009; Pearce, 1997a,b), although work is ongoing to close this gap by analyzing biological approaches to engineering problems in order to obtain novel solutions. Artificial sensing systems are interesting for many applications (Nagle et al., 1998), such as the following:

- healthcare — such as detection of skin or lung cancers,

- quality control — e.g. in agriculture,
- safety — such as fire–sensors, and
- security — such as detection of explosives.

A curious example for commercial application in olfaction, which has not reached market acceptance, are *smellies*, films with accompanying olfactory stimuli. Smellies are possibly still best-known under the Smell-O-Vision brand that came out with the slogan "first (1893) they moved, then (1927) they talked, now (1959) they smell." These efforts failed, however, to become successful for various reasons (Gilbert, 2008, ch. 8) that have to do with technical difficulties and marketing.

In retail settings, smell dispensers have become popular, especially, since they have been shown to stimulate good mood (Baron, 1997). Additionally, congruence with settings can make people explore more, come back more frequently, and stimulate sales (Spangenberg and Grohmann, 2005; Spangenberg et al., 2006).

5.3 Biology

In vertebrates, the early part of the main olfactory system is composed of an olfactory organ, which is the olfactory epithelium, the olfactory nerve, and the olfactory bulb. It can be characterized by several essential conceptual features which are summarized in figures 5.1 and 5.2. They can be stated as the following (after Cleland, 2008):

- Olfactory receptor neurons (ORNs) in the olfactory epithelium (OE) receive information about odorants from their receptors.
- Their axons branch and terminate within glomeruli in the olfactory bulb (OB), where
- they synapse with mitral and tufted cells (commonly called M/T cells).
- These M/T cells transmit information to higher olfactory centers.

Table 5.1 should help to understand convergence and divergence principles in the early olfactory system. Please, note that specific estimates must be treated with caution because of large variability across different studies. This variability is highest for rat GLs and presumably comes from different sources including methods

Figure 5.1: ***Schematic of the early olfactory system*** *including the olfactory epithelium and bulb. Odorant molecules are generally light and lipophilic (Ohloff, 1986). They bind to ORs, which activate ORNs, which in turn transduce the input signal into action potentials. Each olfactory receptor neuron (ORN) generally expresses one type of receptor (OR). This is referred to as the one neuron–one receptor rule (Strotmann et al., 2000; Mombaerts, 2004; Serizawa et al., 2004). Each ORN responds to a set of odorants. Glomeruli (GLs) receive in general only input from ORNs of one type of OR (Mombaerts, 2004; Bozza et al., 2002). This means that each GL is a convergence region for a certain type of receptor with the possible function of signal–to–noise improvement (e.g. Chen and Shepherd, 2005; Pearce et al., 2001; Meisami, 1989). GLs are innervated by the primary dendrites of mitral and tufted cells (M/T cells). These neurons forward their output to higher-order brain regions. Granule cells (not shown) mediate lateral and feedback inhibition on M/T cells. They get excited by the olfactory cortex and M/T cells (Haberly and Price, 1977). The colors and symbols in the plot indicate molecular features of odorants and OR identity of sensory projections to GLs. This figure is taken from Auffarth et al. (2011c, ; paper VI).*

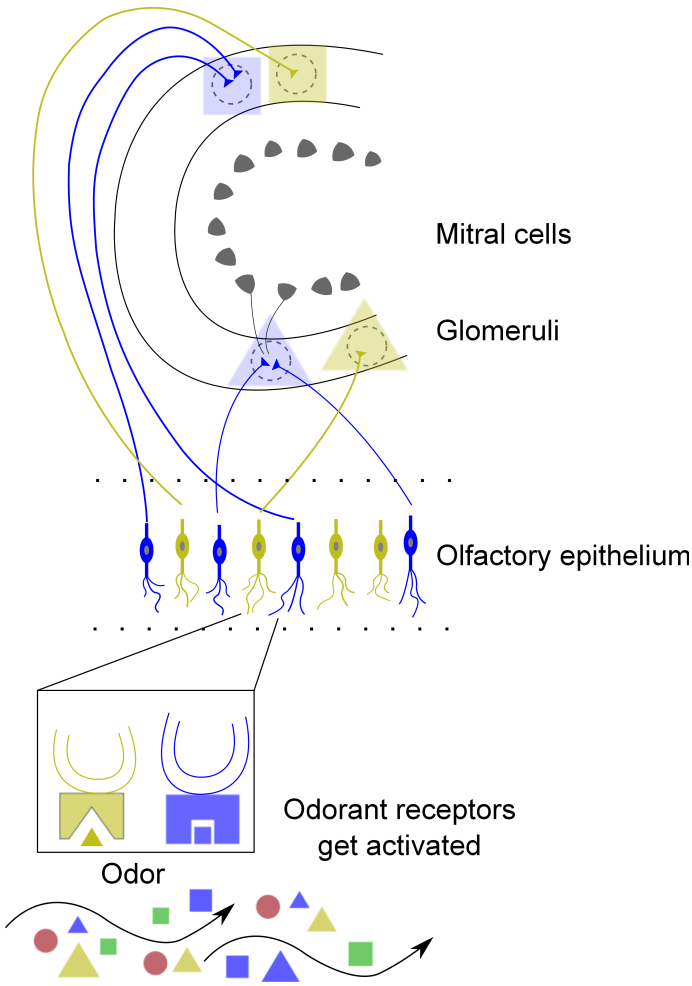


Table 5.1: ***Numbers of functional receptors, glomeruli, and secondary neurons.*** *The table gives estimates for olfactory receptors, glomeruli (GL), and mitral cells per bulb. Mitral cell counts may include some tufted cells. In the* ***rat****, the data are adapted from Zhang et al. (2007b); Royet et al. (1998, 1988). In rats, there is an estimated number of 10 million olfactory receptor neurons (ORNs) per nasal half (Meisami, 1989) and 1.04 million neurons per hemisphere in the piriform cortex (Chen and Buckmaster, 2005). Kim et al. (2011), in their study on the rat, estimated the number of cells in the granule layer (granule cells, deep short axon cells, and possibly others) at roughly between 5,000,000 and 6,000,000. The estimates for the* ***rabbit*** *and the* ***mouse*** *are based on Royet et al. (1998); Meisami and Safari (1981). In the mouse, per glomerulus, there is an estimated number of 20–25 mitral cells, 50 tufted cells, and 25 periglomerular cells (Christie et al., 2005; Purves et al., 2004; Royet et al., 1998).* ***Human*** *data are based on Aloni et al. (2006); Maresh et al. (2008); Bhatnagar et al. (1987). The rounded estimate of M/T cells is for humans at age 25. For comparison, two insect species were added.* ***Honeybee*** *data are based on Robertson and Wanner (2006); Sachse et al. (1999); Gascuel and Masson (1991). Esslen and Kaissling (1976) estimated the number of ORNs in honeybees as 60,000 and of the local interneurons (LIN) as 4,000. Data for the* ***drosophila melanogaster*** *(fruitfly) come from Hallem and Carlson (2004); Vosshall et al. (2000); Clyne et al. (1999); Stocker (1994). Hallem and Carlson (2004) give the number of ORNs in drosophila as 1300.*

	ORs	**GLs**	**mitral cells**
rats	1,300	2350-4,200	56,200
rabbit	?	1,800-6,300	59,600
mouse	1,100	1,810	38,400
human	350-400	5,500	50,000
	ORs	**GLs**	**PNs**
honeybee	163	160	800
drosophila	62	50	150

of measurement and differences in animal care and age. For example, it is known that laboratory animals typically have smaller brains than free animals, which could be partly due to sensory deprivation in laboratory settings (Kempermann et al., 2010) and this is what is observed with the number of GLs in deprived animals (Royet et al., 1998). Bhatnagar et al. (1987) found that volume and cell count of the OB in humans decreases as a function of age. The ratio of glomeruli to mitral cells is likely to vary widely across species (Royet et al., 1998). Royet et al. (1998) suggested

based on their own studies and comparisons to other studies that the number of GLs is often underestimated.

Please refer to fig. 5.1 for more details about the olfactory bulb architecture. There are other cells than M/T cells in early olfaction, including some with supporting functions. Of those involved in bulbar processing, main distinctions are periglomerular cells, short–axon cells, and granule cells (Shepherd et al., 2004).

5.3.1 The peripheral system

In terrestrial animals, odorants are light, volatile molecules that pass through the nasal passages (called orthonasal stimulation) and via the mouth (retronasal stimulation)[2]. On the olfactory epithelium, towards the dorsal roof of the nasal cavity, these molecules dissolve in the mucosa, an aqueous medium, and bind with receptors (Buck and Axel, 1991). These are located on hair–like projections, called cilia, on top of receptor cell protrusions (dendritic knobs; cf. fig. 5.3).

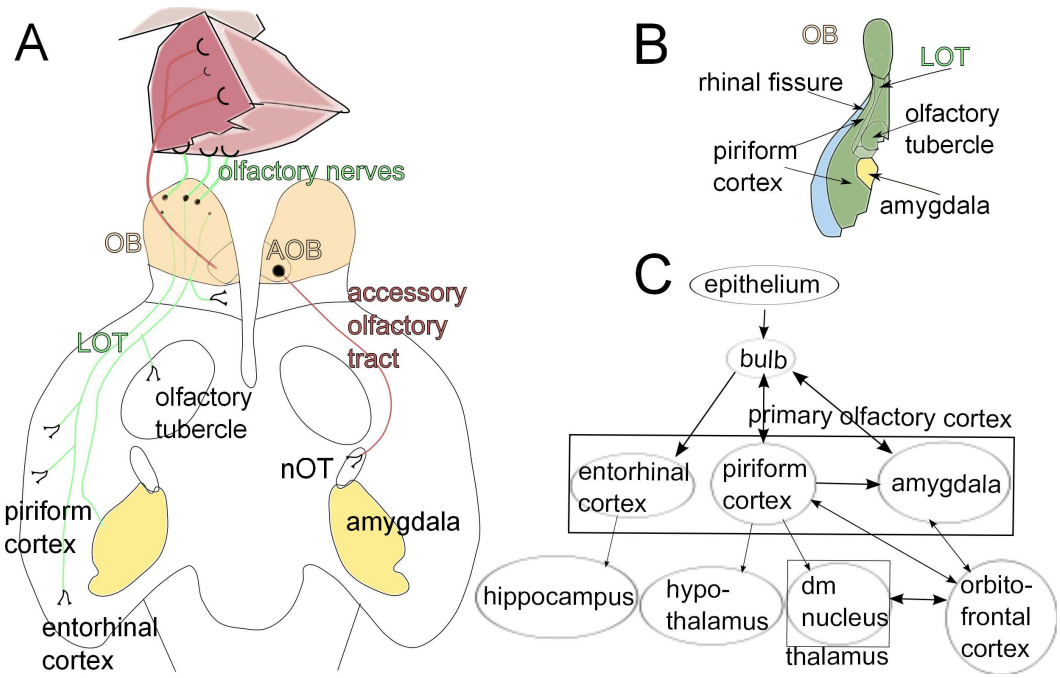
In contrast to pheromone receptors, which only respond to particular pheromones, odorant receptors have been characterized as *broadly tuned* (c.f. Malnic et al., 1999; Sánchez-Montañés and Pearce, 2002; Wilson and Mainen, 2006; Nara et al., 2011). However, Nara et al. (2011) recently found in a large–scale analysis of olfactory receptor neurons that most receptors respond only to few odorants.

Generally, it is believed that the recognition of odorants by receptors is an analytical process (however c.f. Turin, 1996). *Odotope theory* (Mori, 1995; Mori and Shepherd, 1994; Shepherd, 1987) is the prevalent view on olfactory transduction and proposes that receptors are molecular feature detectors, although it is not clear what these features are. These features are called *odotopes* in analogy with epitopes, the antigenic determinant of the immune system. Each odorant molecule contains many different properties and according to this theory, the information about the odorant is then encoded by the combined responses of many types of receptors, each of which recognizes a specific subset.

It is believed that the chemosensory repertoire is of perceptual consequence (c.f. Keller and Vosshall, 2008). Keller et al. (2007) found that a genetic variation in

[2] In aquatic animals, odorants are water soluble.

Figure 5.2: ***The rat olfactory system****: Anatomical illustration of olfactory bulb and downstream areas.* ***A*** *— The peripheral input is through axons coming from the olfactory epithelium. This information is then transmitted over the bulb to the primary olfactory cortex (also called piriform cortex).* ***AOB*** *— accessory olfactory bulb.* ***nOT*** *— bed nucleus of the olfactory tract. This figure adapted from Watson et al. (2010, p. 101).* ***B*** *— Anatomical drawing of the inferior aspect of the right half of a rabbit's brain including olfactory bulb and downstream areas. Adapted from Cunningham (1918, p. 625).* ***C*** *— Abstract schematic of areas and information flow in the main olfactory system including the bulb as well as primary olfactory cortex and other downstream areas.* ***dm nucleus*** *— dorsomedial nucleus. Olfactory projections go from the olfactory bulb directly to primary and secondary olfactory areas via afferent fibers, the lateral olfactory tract (****LOT****). The olfactory system is special in the sense that olfactory signals pass to the primary sensory cortex without a thalamic relay. Projections connect to so–called limbic structures such as amygdala and hippocampus. The OB is connected to the amygdala over slow and fast pathways, directly and indirectly over the piriform cortex. The anterior commissure is a large bundle of crossing fibers which connects contralateral OBs and parts of the cerebrum. This figure was constructed based on Shipley et al. (2008), de Castro (2009), and Wilson and Stevenson (2006). Compare also Haberly and Bower (1989) and (more generally about large–scale integration of the piriform cortex) Averbeck and Seo (2008).*



a single OR in humans is related to differences in perception of androstenone and andostadieonone. Khan et al. (2011) reported that the frequency distribution of receptor gene expression is not uniform. There seems also to be a variability of gene expression during development (Nguyen et al., 2010).

In vertebrates, each olfactory receptor neuron (c.f. fig. 5.3) expresses generally one (Strotmann et al., 2000; Mombaerts, 2004) of a possible functional 1,300 ORs (roughly 350 identified functional types in humans — Aloni et al., 2006; Glusman et al., 2001; Zhang et al., 2007a). ORNs are apparently not scattered at random over the olfactory epithelium (Spehr, 2010; Mori, 1999). Gardner et al. (2007); Mozell et al. (1987); Hornung and Mozell (1977); Mozell (1964) suggested that the distribution of ORNs along the turbinates could be optimized in a way analogous to a chromatograph, however, Abaffy and Defazio (2011) found that the spatial order of ORs was not dependent on volatility and solubility of odorants.

ORNs transmit signals about the odor to the olfactory bulb, a multi–layered telencephalic brain structure. (Please, refer to the microscopy image in fig. 5.3.)

5.3.2 The olfactory bulb

The olfactory bulb (OB) is spheroidal structure (c.f. fig. 5.4) in the brain, which is crucial for the processing of olfactory signals (c.f. Buschhüter et al., 2008; Cleland et al., 2007; Johnson and Leon, 2000d). The processing units on its surface are called glomeruli (GLs). These GLs are mostly devoid of cell bodies, built of neuropil and measure approximately $100 - 200\,\mu m$ in diameter (Pinching and Powell, 1971). They are the place where sensory axons synapse with bulbar interneurons and dendrites of OB output neurons (M/T cells; Kosaka and Kosaka, 2011).

Axons from ORNs that express the same OR are bundled in approximately two glomeruli at stereotyped positions (typically one on each of both mirror–symmetric sides) of about 1800 in each olfactory bulb (Fantana, 2006; Zhang et al., 2007b; Ressler et al., 1994; Vassar et al., 1994). Each glomerulus is target for convergence of axons from many ORNs expressing (mostly) the same type of olfactory receptor (OR; Chen and Shepherd, 2005; Bozza et al., 2002; Mori, 1999; Kauer and Cinelli, 1993; Shepherd, 1992). This convergence, on the order of several thousand ORNs per glomerulus (c.f. table 5.1), could allow for reduced detection thresholds and higher

Figure 5.3: ***Anatomical drawings of olfactory receptor neuron and microscope image of the olfactory bulb.*** *(**a**) Illustration of olfactory receptor neurons (ORNs). Olfactory receptor neurons (ORNs) are bipolar, unmyelinated, and very thin* ($0.1 - 0.3$ µm *of diameter; Shepherd et al., 2004). Image source: Cunningham (page 805 1918). (**b**) A coronal section through the main olfactory bulb (OB) of an adult male mouse. Image taken with a confocal microscope visualizing all cell nuclei. The color code indicates the following cells:* ***blue*** *— glomerular layer containing periglomerular cells,* ***red*** *— external plexiform layer (EPL) and mitral cell layer containing cell bodies of M/T cells and some granule cells,* ***green*** *— internal plexiform and granule cell layer containing cell bodies of immature migrating neuroblasts, and mature granule cells. Image source: Valley (2006). (**c**) A sagittal section through the olfactory bulb of an 16 day old mouse. Tissue was stained using immunoperoxidase with antibodies against a* GABA_A*–receptor (brown) and counter stained using Nissl (blue). Indicated are glomeruli (**G**), the mitral cell layer (**MC**), and the accessory olfactory bulb (**AOB**) which receives input from the vomeronasal organ. Image source: Elsaesser and Paysan (2007).*

(a) *Olfactory receptor neuron*

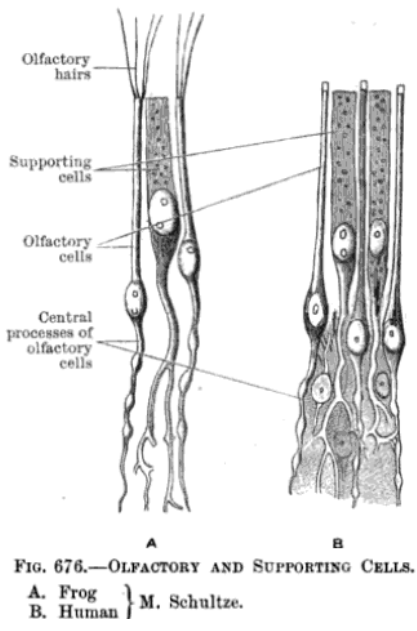
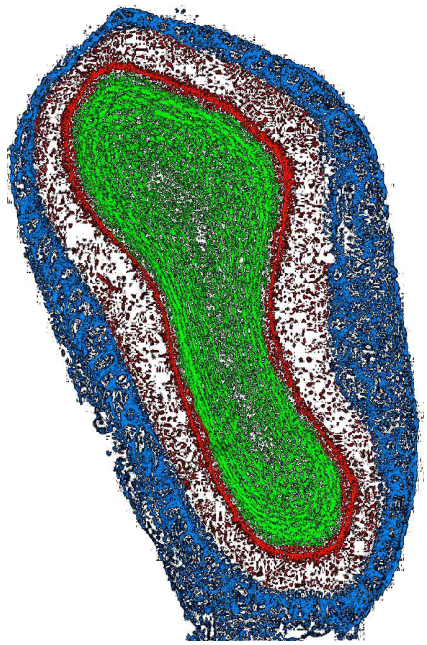


FIG. 676.—OLFACTORY AND SUPPORTING CELLS.
A. Frog
B. Human
} M. Schultze.

(b) *Olfactory bulb (coronal section)*



(c) *Olfactory bulb (sagittal section)*

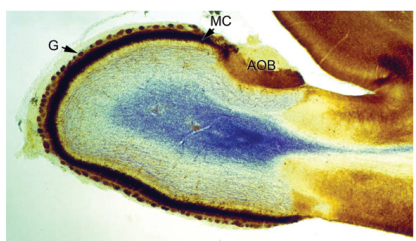
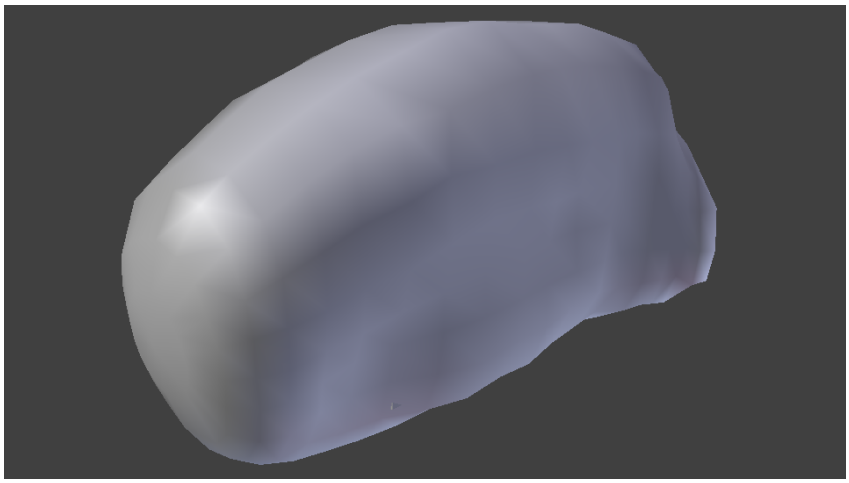


Figure 5.4: ***Rendered 3-D model of the olfactory bulb.*** *The olfactory bulb is a spheroidal structure. This image was generated from a Open–GL rendering using the Blender software with a vertex model provided by the Leon lab (Leon and Johnson, 2006).*



sensitivity which also implies broader dose–response curves (Chen and Shepherd, 2005; Meisami, 1989; Pearce et al., 2001).

To summarize the basic principles of this early olfactory processing:

- Every ORN expresses a single type of olfactory receptor (one neuron–one receptor rule)[3];
- Each OR responds to many molecules and each molecule elicits responses from many different receptors;
- ORN axons are targeting GLs;
- Each glomerulus receives input from neurons of (generally) only one type of receptor (one glomerulus–one receptor type rule).

In this fashion, each odorant elicits a specific pattern of glomerular activation.

[3] An exception from this rule is the Caenorhabditis elegans

5.3.3 Axon guidance

The establishment of a functional architecture in networks of neurons during development and its maintenance in mature animals relies on mechanisms of process outgrowth and synapse formation. This formation of connectivity is defined by several factors, including molecular cues and neural activity (Butz et al., 2009; Katz and Crowley, 2002; Sperry, 1963), but also has been linked to the exploration of the environment (Laishram et al., 2009) and mechanical factors (Cardamone et al., 2011; Weiss and Hiscoe, 1948). Sensory axon coalescence has been found to rely on several mechanisms that contribute differently on a local and global scale, some of which are likely related to activity (Mori and Sakano, 2011; Sakano, 2010; Mombaerts, 2006; Ming et al., 2002). It is known that ORN type–convergence is at least partly mediated by experience (Imai and Sakano, 2007; Kerr and Belluscio, 2006; Yu et al., 2004). Particularly, in the absence of activity cues, GLs can contain projections from ORNs of different OR types (Zou et al., 2004).

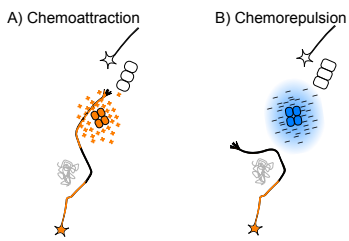
In axonal growth, the direction of the growth cone is regulated by various chemical cues, diffusible chemo–attractants and repellants (Sanes and Jessell, 2000). The fasciculation and targeting of axons are promoted by the expression of receptors for guidance and adhesion molecules, which can be regulated by neuronal activity (Cho et al., 2009). Serizawa et al. (2006) found evidence in the mouse that a correlation of neural activity mediated axonal attraction and repulsion by up– and down–regulation of a set of olfactory axon guidance cues. This indicates that axon sorting could be based on correlated neural activity. The principles of chemo–attraction and –repulsion in axon fate are illustrated in fig. 5.5.

It is known that in vision, touch, sound, and taste, a topographic order of representations is maintained in different regions of the brain (c.f. section 6.4.1). This topography could come about by several principles, including by the aforementioned molecular gradient information (McLaughlin and O'Leary, 2005). In a pioneering experiments in frogs, after dissection of the retino–tectal connections and surgical rotation of the retina, the connections regrow and result in a rotated topography, with behavioral consequences such as tongue flicking responses in opposite directions (Sperry, 1944, 1945).

Topography leads to a grouping of units with similar response profiles. Therefore,

self–organized models have been proposed as a model of map formation (e.g. Antolík and Bednar, 2011; Tsigankov and Koulakov, 2010; Swindale, 1996; Kohonen, 1982; Willshaw and Malsburg, 1976; Malsburg, 1973), which have been demonstrated to result in a topographic arrangement. Biologically, this correlation principle could be based on global or local distribution of gradients of cues in axonal guidance (Serizawa et al., 2006; Ming and Song, 2005). Dubacq et al. (2009) demonstrated recently a local synthesis of the receptor mRNAs in the sensory axons, a mechanism which could result in a sorting based on OR identity.

Figure 5.5: ***Chemoattraction and chemorepulsion in axon guidance****. Principles of axon guidance through chemoattraction and chemorepulsion: Direction of the axon's growth cone is regulated by various chemical cues in a series of discrete steps Sanes and Jessell (2000). The axon in the middle (**light orange**) receives directional information from diffusible repellants (**blue**) and from attractive guidance molecules drawing the axon in direction of the left axon (**orange**). The fasciculation and targeting of ORN axons is promoted by the expression of adhesion molecules and other guidance molecules, which can be regulated by neuronal activity. ORN axons can express different receptors for guidance molecules. The combination of the receptors expressed may be related to the strength of adhesive and repulsive forces between axons.(Cho et al., 2009). This figure was created by Bernhard Kaplan.*



5.3.4 Bulbar processing

The olfactory bulb circuit contains axodendritic, dendrodendritic, somatodendritic, and dendrosomatic synapses (Shepherd et al., 2004). Apart from periglomerular, short–axon, and granule cells, many other cell types have been identified and are still being identified (c.f. Kosaka and Kosaka, 2011).

Willhite et al. (2006) documented evidence from a tracer study with virus injection for a columnar organization by OR identity. These columns stretched across glomerular, mitral cell, and granule cell layers. Mori et al. (2009) called this functional compartmentalization *receptor channels*.

5.3.4.1 Interglomerular circuitry

Yaksi and Wilson (2010) provide evidence in the insect that local circuitry between GLs could serve for gain control (also called *volume control*) by contrast enhancement. The length of the periglomerular axons, which strech only across few GLs (Shepherd et al., 2004) could speak in favor of this functionality. Olsen and Wilson (2008) demonstrated in drosophila that lateral inhibitory activity between glomeruli scales with input activity and therefore serves for concentration normalization (for gain control in the locust also compare Papadopoulou et al. 2011; Luo et al. 2010). It could also be that concentration invariance is learned as part of the wiring plasticity between glomeruli (Cleland et al., 2012). Gill and Pearce (2003) compared learning rules for wiring models between GLs.

5.3.4.2 Granule cells

Granule cells receive input from M/T cells and the olfactory cortex. They laterally inhibit mitral cell activities via dendrodendritic synapses (c.f. review in Shepherd et al., 2007;Price and Powell, 1970). Willhite et al. (2006) found in a viral tracer study that granule cells receive feed–forward input from mainly one glomerulus. In another tracer study, they found that inhibitory connectivity onto M/T cells was mediated by a broad, spatially segregated subset of granule cells (Kim et al., 2011). They speculated that this connectivity pattern could help to shape the OB output to simplify the readout of behaviorally relevant information.

M/T cell lateral inhibition is thought to serve a computational function that has been called olfactory contrast enhancement (Abraham et al., 2010; Yokoi, 1995). Migliore and Shepherd (2008) and Wiechert et al. (2010) developed computational models of this circuit functionality.

Granule cells receive axosomatic and axodendritic input from the olfactory cortex. This input introduces modulation of odor responses by high–level processes

such as attention, expectancy, and memory (Freeman and Schneider, 1982; Kay and Laurent, 1999). Additionally granule cells receive input from the anterior commissure of the contralateral OB.

Pager et al. (1972) compared mitral cell responses to food odors by bipolar electrodes and found a response increase in hungry versus satiated rats, which they discussed in terms of the bulbar–hypothalamic loop. Lledo and Lagier (2006) suggested that plastic mechanisms in the OB are highly modulated and can serve to optimize performance for a given environment. Similarly, Fuentes et al. (2008) measured single M/T cell responses and local field potentials in behaving awake rats. They found a reduction of firing rates and a decrease in power at frequencies above 50 Hz during a two–alternative choice discrimination task versus passive odorant exposure which they attributed to task–dependent modulation by top–down projections. On the other hand, Li et al. (2011) found that there was no significant difference in mitral cell activities when the odorant was presented at different anesthesia levels. They argued for existence of mechanisms in the OB that ensure that odor information reaches the sensory cortex relatively unmodulated by centrifugal connections.

5.3.4.3 Output neurons

Mitral and tufted cells are often classified together as M/T cells (such as in this thesis; c.f. Mori, 1999), although tufted cells are smaller and comparably closer to the GLs, and they appear to have sometimes different response properties and projection patterns (c.f. Nagayama et al., 2010, 2004). Nagayama et al. (2010) presented data suggesting that mitral cells projected rather to anterior and posterior piriform cortex, while tufted cells projected to the olfactory tubercle.

While Willhite et al. (2006) found that M/T cells receive feed–forward input from mainly one glomerulus, Fantana et al. (2008) found that mitral cells receive effective inputs from few, chemically dissimilar, and spatially separated GLs. This apparent contradiction could be reconciled by findings by Najac et al. (2011), who demonstrated a dendrodendritic excitation between M/T cells over their secondary dendrites (which occurs on a slower timescale than postsynaptic excitation by ORNs). Overall, M/T cells seem to have a more narrow response profile than glomeruli (Tan

et al., 2010). Such a circuitry could point towards a role of M/T cells as feature detectors.

5.3.5 Symmetry and crossings

In the visual system, axons from ganglion cells in the medial retina cross in the optic chiasm to the contralateral side and similar symmetry has been observed in the olfactory system, although this seems to be a less prominent feature (as discussed in Huesa et al., 2000). In the olfactory system, there is symmetry between the two bulbs (c.f. Rubin and Katz, 1999), as well as a symmetry within the olfactory bulb.

Sensory axons coming from a neuron of the same receptor type commonly project to about two GLs on each side, the positions of which are symmetric along a radial axis (c.f. Khan et al., 2010; Soucy et al., 2009; Lodovichi et al., 2003; Vassar et al., 1994). Lodovichi et al. (2003) reported that external tufted cells form intrabulbar connections to granule cells in the isofunctional olfactory columns.

There is evidence in some animals for primary axons crossing the midline over to the contralateral olfactory bulb. Secondary fibers cross over the anterior commisure from mitral cells of one side to granule cells of the contralateral side. Yan et al. (2008) revealed in a tracing study that isofunctional olfactory columns are connected between the two bulbs over excitatory neurons in the anterior olfactory nucleus pars externa (connections running via the anterior commissure).

In fact, it was experimentally shown that human subjects can recognize odors over the naïve nostril that was not exposed during previous learning (Mainland et al., 2002), although this result could be explained by centrifugal connectivity, as well.

In insects, there is a bilateral symmetry between the two antennal lobes, (for example Galizia et al., 1998; Fiala et al., 2002). In contrast to vertebrates, to our knowledge, an internal symmetry has not been reported.

5.3.6 Downstream processing

All brain regions that receive direct bulbar input are collectively called *olfactory cortex* (de Olmos et al., 1978; Price, 1973). Principal recipients include the piriform cortex, the amygdala, and the entorhinal cortex. Olfactory bulb output neurons,

M/T cells, project to primary olfactory cortical areas, anterior olfactory nucleus, piriform cortex, olfactory tubercle and lateral entorhinal cortex, and the amygdala (Shipley et al., 2008, ; c.f. fig. 5.2). Most of the olfactory cortex is of an evolutionarily old 3–layered type.

The olfactory cortex is the only area of the cortex that receives direct sensory input without intermittent thalamic connection. While this could be understood to mean that a thalamic relay is inconsequential for smell or that the piriform cortex plays the role of the thalamus, results from Plailly et al. (2008) suggest that olfaction, like all other sensory modalities, requires the thalamus when consciously analyzing.

The main task of the olfactory cortex was suggested to be synthesis of different dimensions into perceptual objects (c.f. Haberly, 2001; Wilson and Stevenson, 2006; Gottfried, 2010). Savic et al. (2000) found evidence that response–activity to smells in higher–order brain centers is task–dependent, which suggests a parallel and hierarchical processing of olfactory information (c.f. also Savic and Berglund, 2004).

Anderson et al. (2003) found in an fMRI study with human subjects that amygdala activation was related to intensity but not valence (aversive vs. pleasant) of odors. Further, they found activations in the orbitofrontal cortex to be related to valence. Grabenhorst et al. (2007) found in another fMRI study that pleasant odors lead to an increase of activations in the medial orbitofrontal cortex, while unpleasant odors activate differentially the cingulate and mid–orbitofrontal cortex. Rolls (2011) proposed that the orbitofrontal cortex is associated with reward value of stimuli (visual, olfactory, and taste).

Several studies have looked at convergence of M/T cells to the olfactory cortex (Miyamichi et al., 2011; Ghosh et al., 2011; Sosulski et al., 2011; Gottfried, 2010; Miyasaka et al., 2009; Stettler and Axel, 2009). Using different tracing techniques in the mouse, Miyamichi et al. (2011); Ghosh et al. (2011); Sosulski et al. (2011) found no spatial bias in connectivity of M/T cells to the piriform cortex. This suggests that M/T cells which innervate the same glomerulus, independently connect to postsynaptic cortical targets in the piriform cortex. This is different from the situation in other sensory cortices, compare section 6.4.1. Currently, the consensus is that cells in the piriform cortex are fed by convergent input from random collections

of GLs without any order in the spatial domain (cf. Choi et al., 2011).[4] If confirmed, this apparent non–topographic projection could indicate that apart from complex higher–order feature selectivity there is a reorganization of the sensory space in the olfactory cortex.

It is known that representations in the human piriform cortex are distinct for several olfactory perceptual qualities (Gottfried, 2010). Howard et al. (2009) showed that fMRI measures in the posterior piriform cortex in humans (but not of the anterior piriform cortex, amygdala, or orbitofrontal cortex) correlated with perceptual differences. Choi et al. (2011) showed that populations of neurons in the piriform cortex can drive behavior. In yet another fMRI study, Gottfried et al. (2006) suggested based on experimental data, again from humans, a bimodal processing of odors with a dissociation in the piriform cortex between odorant structure and odor quality such that anterior regions encode structure and posterior regions encode quality. Davison and Ehlers (2011) found by in–vivo electrophysiological recordings in mice the anterior piriform cortex that firing was insensitive to single–glomerulus photostimulation, but instead showed responses to combinations of coactive GLs, which supports their proposed role as higher–order feature detectors. Yoshida and Mori (2007) compared response profiles of neurons in the dorsoposterior part of the rat piriform cortex across multiple food categories and found that neurons were selective to either one or a combination of several of these categories, from which they concluded that neurons in the piriform cortex have a category–profile selectivity.

Miyamichi et al. (2011) did find patterned connectivity of M/T cells to the amygdala and anterior olfactory nucleus (AON).

We will return to this in chapter 6, where we will talk about representations in the olfactory system.

[4] Zou et al. (2008, 2005) reported in mice that different zones in the mucosa project to different parts of the bulb and that in the piriform cortex, the areas of activation overlapped, which suggests that complex odors are processed cortically. However, they later failed to replicate their results and Linda Buck retracted their articles.

5.4 Understanding smell

Odor–related spaces are illustrated in fig. 5.6. They correspond to genetics, the olfactory stimulus, neurophysiology, and cognition, respectively. In case of vision, the retina contains light–sensitive rods and cones. Color vision is mediated by separate classes of cones each tuned to different frequencies of light. In trichromatic mammals, three types of cones correspond to the three basic colors, red, green, and blue. In audition, organs in the inner ear distinguish between resonance frequencies. In taste, five basic tastes are sensed by taste receptor cells which have their specific sensitivities to a single taste quality. In most vertebrates, there exist roughly between 600 and 1300 OR genes, with a large fraction being pseudogenes — the exact numbers of genes and the fraction of pseudogenes vary considerably between species (c.f. Kaupp, 2010). The transduction principle of receptors is not well–understood (for a recent introduction see Crasto, 2009) and the physico–chemical space is possibly highly complex (Haddad et al., 2008b). Each odorant triggers one of a possibly indeterminate number of activation patterns on the epithelium, the OB, and downstream layers. These patterns of activities are perceptually interpreted.

Elucidating the interactions and links between relationships between these different spaces could be pivotal in understanding olfaction. In order to accomplish this different approaches are necessary such as investigating intermediate representations, influences of the behavioral context, and computational principles.

5.4.1 The stimulus problem

The broad availability of gas chromatography (James and Martin, 1952) in conjunction with mass spectroscopy allowed for the chemical analysis of natural odors (for example Friedrich and Acree, 1998) and, using an olfactometer, a device for delivery of gases, which allows for control over mixtures, humidity, and temperature, studies of perceptual qualities of odors (Zwaardemaker, 1917) could be done (see Reineccius and Vickers, 2000 for an introduction to perceptual and chemical analysis of olfactory, gustatory, and trigeminal stimuli). Some complex odors, such as wine or coffee, are composed of hundreds of different mixture components (Briandet et al., 1996; Louw et al., 2009). Which of these components and which of the physico–chemical

properties are important, presented itself as a problem. It turns out that many of these complex smells can be reduced to the impact of a few molecules (Gilbert, 2008, p. 28), so, for example, it was found out that the main contributors to fecal smell are sulfur–containing compounds, and compounds including methyl benzoate (Woodford, 1981) are now used to train dogs to detect cocaine, and smell of Gala Apples corresponds to L-Alanine, described by the Hill formula $C_{136}H_{215}N_{33}O_{45}$ (Sitzmann, 2009).

The *stimulus problem* in olfaction, which refers to the relation of physical properties and perception which is poorly understood (c.f. Sell, 2006; Turin, 2002; Rossiter, 1996; Ohloff, 1986). Alexander Bell is quoted as having said in 1914 (Wise et al., 2000):

> Can you measure the difference between one kind of smell and another? It is very obvious that we have very many different kinds of smells, all the way from the odour of violets and roses up to asafetida. But until you can measure their likeness and differences you can have no science of odour.

Ninety years later, in his review "On the Unpredictability of Odor," after reviewing research, Charles Sell concluded that there are no molecular features of the odorant that directly determine perceptive quality and that it "would seem that consistently accurate prediction of odors are not possible for a very considerable time" (Sell, 2006). Aggravating this complicated situation is that there are interactions between features (for example Johnson et al., 2005a) and molecules in a mixture (Zwaardemaker, 1917).

Turin discusses the stimulus problem at length in his book *The secret of scent: adventures in perfume and the science of smell* (Turin, 2006). Some odorants of similar structure smell similar, such as guaiacol and vanillic, however, different structure sometimes comes also with similar smell such as benzaldehyde and cyanide, which smell of bitter almonds. Further, a similar structure can come with a different smell, such as straight–chain aldehydes. He puts up hypotheses to account for this, including: (i) different receptors for the smell and (ii) the molecules get associated together because they often come together.

Size and volatility of odorants affect their usefulness in communication (c.f. Wilson and Stevenson, 2006, chapter 5). Small volatile molecules allow for rapid diffusion over short distances, while large and less volatile molecules provide a relatively long signal. At the same time, structural complexity (including length and weight) of odorants is related to the complexity of the smells and their pleasantness (Kermen et al., 2011; Zarzo, 2011; Joussain et al., 2011).

Haddad et al. (2008a) proposed a space of 32 dimensions of molecular properties that co–varied with activation patterns of glomeruli and MT cells. However, it can be expected that the last word on this topic is not spoken yet. Soh et al. (2011) found another metric based on structural similarity of odorants that seemed to outperform Haddad and colleagues measure on one dataset, while performing on a similar level on another. Chen et al. (2011b) found only barely significant correlations of this metric with results from a behavioral study on drosophila. Also, as noted by Chen et al. (2011b), Haddad et al. (2008a) did not account for effects of intensity.

It is thought that molecular features relate mostly to molecular shape and one physico–chemical dimension that investigations have put emphasis on is carbon chain length of odorants (e.g. Mori et al., 2006b; Leon and Johnson, 2003; Johnson and Leon, 2000d). It is known, however, that enantiomers, molecules that are mirror–symmetric to each other, can produce different activations at the bulb and cortical levels and different odor sensations (Leon and Johnson, 2003; Li et al., 2008), which indicates that shape alone does not account for relevant features.

It has been found that principal dimensions in a dataset of odorants described by physico–chemical properties are related to pleasantness (Khan et al., 2007). Khan et al. (2007); Haddad et al. (2008b) argued that olfactory pleasantness corresponds to a natural axis of maximal discriminability among biologically relevant molecules and that the olfactory system has evolved to exploit regularities in the odor space.

There exist many studies on perception of odor intensity and detection thresholds (e.g. Kobal and Kettenmann, 2000; Cain and Johnson, 1978; Zwaardemaker, 1909), usually done using an olfactometer. Laska and Teubner (1999) found in a forced–choice test that the discrimination ability of subjects between homologous odors was correlated to differences in carbon chain length. It is known that some odors are associated with specific molecular properties, e.g. putrid to amines and Doleman

(1998) suggested that increased olfactory sensitivity for alkylamines or alkylthiols (in humans) as compared to alkanes or alcohols could be accounted for by evolutionary adaptation for detecting decaying food and toxic gases.

In section 6.4.3 the importance of different molecular features will be discussed in the context of OB representations.

5.4.2 Perceptual dimensions of smell

The epistemological status of olfaction was disputed during a long time Le Guérer (2002) and therefore it is not surprising that literature on olfactory dimensions is not abundant (however, see review in Zarzo and Stanton, 2009, and references therein).

Some authors speculated about a number of olfactory perceptions, e.g. Axel (2006) put this number up to 10,000, however, there is little evidence to back this up and coding mechanisms in the epithelium (Malnic et al., 1999) and the bulb (Meister and Bonhoeffer, 2001; Sachse et al., 1999; Spors and Grinvald, 2002; Uchida et al., 2000; Wachowiak and Cohen, 2003, 2001; Kauer and Cinelli, 1993) suggest rather an unlimited number.

Generally, perceptual characters of odors have been described by two differrent methods (Zarzo and Stanton, 2009; Wise et al., 2000; Chastrette, 1998): (i) numerical scales of odor similarity assigned by a panel or (ii) semantic labeling. Particularly well–known examples are, for the first type, Dravnieks' panel assessment of odor character (Dravnieks, 1985), and, for the second type, semantic databases such as Acree and Arn (1998). Similarly to the case in physico–chemical properties, the dimensionality of these datasets is quite high. Callegari et al. (1997) analyzed correlations between odorants as a function of the number of descriptors and argued that around 30 descriptors are optimal. Analyses on these datasets commonly observe an embedded structure of latent dimensions of categories which explain the perceptual experience or the common effect of odorants. One of the most salient dimensions in these data seems to be pleasantness (Arzi and Sobel, 2011; Yeshurun and Sobel, 2010; Zarzo, 2008). Generally, it is plausible that things that are potentially useful for the body tend to smell good, and things that are potentially harmful smell bad[5].

[5]This rule has some exceptions, such as some food which actually smells unpleasant to people

Different categorizations have been proposed for natural odors (reviewed in Gilbert, 2008, , p. 5–37; Chastrette et al., 1988; Abe et al., 1990; Mamlouk and Martinetz, 2004; Zarzo and Stanton, 2006). They include the wine aroma wheel (Noble et al., 1984), a beer flavor wheel (Meilgaard, 1982), for medical plants (Linneus, 1752), or more general for all kind of smells (Crocker and Henderson, 1927). Amoore (1977) proposed basic categories from data on anosmia, the loss of smell. These are listed below with examples:

- Camphoraceous (such as mothballs)
- Musky (perfumes and aftershave),
- Floral (roses),
- Pepperminty,
- Ethereal (cleaning fluid),
- Pungent (vinegar),
- Putrid (rotten eggs).

It is known that the categorization of odors varies with cultural differences (Chrea et al., 2005; Chrea, 2004; Ayabe-Kanamura et al., 1998), however Chrea et al. (2005) argued that odor categories are based on perceptual similarities and that, while the category boundaries are culture–specific, core odor–category structure could have common representations. Zarzo and Stanton (2009) found that the PCA space of physico–chemical odorant descriptors from several studies shows a clustering by odorant categories, which could suggest that some perceptual categories are more universal than previously thought. As they reconciled results from different datasets, they found proximities between descriptors such as a group of sweet, balsam, vanilla, heliotropin, and syrup (as reviewed by Donna, 2009).

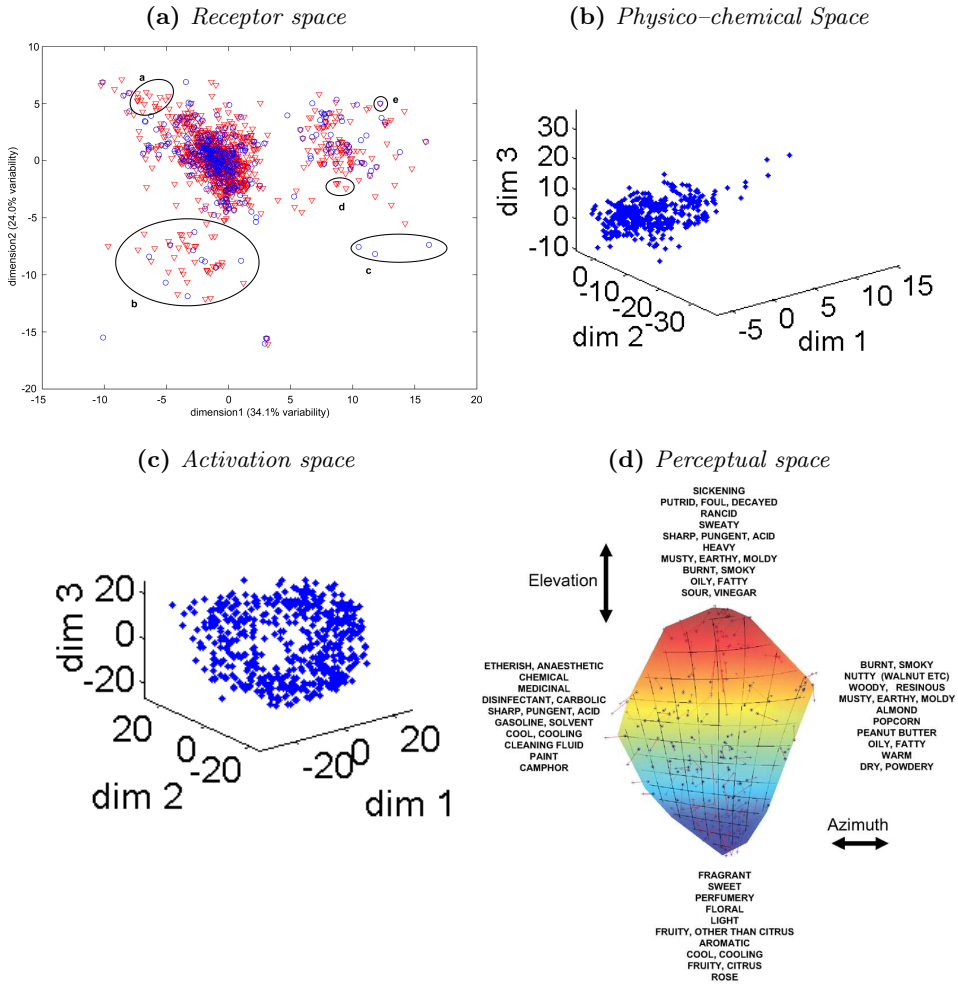
This structure of the perceptual space could be partly grounded in physico–chemical properties. Khan et al. (2007) found by principal component analysis of physico–chemical descriptors that latent dimensionality in physical space is related to hedonic quality of odorants. Mandairon et al. (2009) analyzed investigation time in mice with different odorants and found that they were related with human

not used to it, such as some cheese, coffee, fermented herring, etc.

pleasantness sensation of these odors. They argued that perception of odors in mice and humans is related and speculated that this similarity could arise because of prewiring from receptor level.

Smell is not an intrinsic property of odorant molecules, but a perceptual phenomenon that depends on mechanisms proper to the biological organism which perceives. Studying activations elicited by odorants can help to understand how smells are generated. The next chapter will deal with representations in the olfactory system, especially at the glomerular layer.

Figure 5.6: ***Spaces of smell.*** ***A*** *— Isomap projection of receptors similarity based on distances between genetic coding regions of predicted binding pockets. Circles a–d indicate regions where binding sites of either human or mouse are over–represented. Circle e indicates a pair of human–mouse orthologs (MOR27-1 and OR52P1) with identical binding site. Source: Man et al. (2007).* ***B*** *— MDS projection onto three dimensions of the 32–dimensional matrix of molecular properties proposed by Haddad et al. (2008a). Each point represents an odorant.* ***C*** *— Euclidean differences between glomerular odorant response patterns, computed based on data from Leon and Johnson (2006).* ***D*** *— MDS projection of odorant perceptual descriptors. Source: Koulakov (2011).*



Chapter 6

Representation of odors

In the previous chapter, it was briefly discussed how odorants, which generally encompass many air–borne mono–molecular components, are analyzed into component features and how this information is propagated via electrical activity to the OB. This chapter will discuss how odors are expressed in terms of spatio–temporal patterns of neural activity at the level of the OB, specifically at the glomerular level. How sensory objects are generated from molecules in the air is not clear, but as we will see, such patterns are likely the basis for perceptual processes. Where this happens is also unclear, however, a perceptual ordering of representations has been found previously in the epithelium (Lapid et al., 2011). Others have found such an order in the piriform cortex (Howard et al., 2009) and, recently, we found an order by perceptual categories in the OB (Auffarth et al., 2011c; paper V).

Many studies of odorant coding in the OB provide evidence that odors are represented at the levels of glomeruli and M/T cells by spatio–temporal codes (c.f. Niessing and Friedrich, 2010; Leon and Johnson, 2009; Rubin and Cleland, 2006; Lei et al., 2004; Spors and Grinvald, 2002; Laurent, 1997), while there is uncertainty with regard to several questions, which will be briefly introduced in the following sections:

- population vs. spatial codes (c.f. section 6.3 and section 6.4),
  - distributed vs. localized (clustered) representations (c.f. section 6.4.1),
- the relative importance of spatial versus temporal patterning (c.f. section 6.2).

In the case of local spatial representations, another question pertains to the principles of distance relationships between representations (c.f. section 6.4.3).

Generally, the topology of information processing in brain regions is constrained by its task demands (Sporns, 2010; Montagnini and Treves, 2003), which includes its input–output relationships with other regions. While topographic organization (local and spatial) indicates bottom–up integration or coincidence detection of mixed information, distributed codes or population codes would indicate a wide broadcast of information and often is not explained by purely bottom–up approaches (Cauller, 1995).

6.1 Neuroimaging methods for the OB

There are many different techniques for visualizing processes in glomeruli (for reviews see Galizia, 2009; Pain et al., 2011), and the usual caveats apply for the use of signals with respect to their origin, time, and spatial resolution (c.f. fig. 6.1; for a review with many techniques relevant to olfaction, especially in insects, see Galizia, 2009). Various different methods of monitoring glomerular activity have been utilized, including voltage–sensitive dyes, optical imaging, calcium imaging, and 2DG (c.f. Manzini et al., 2007; Takahashi et al., 2004; Meister and Bonhoeffer, 2001; Spors and Grinvald, 2002; Wachowiak and Cohen, 2003, 2001; Uchida et al., 2000; Sachse et al., 1999; Friedrich and Korsching, 1997; Stewart et al., 1979).

Please, refer to section 7.1 for more details on the 2DG method.

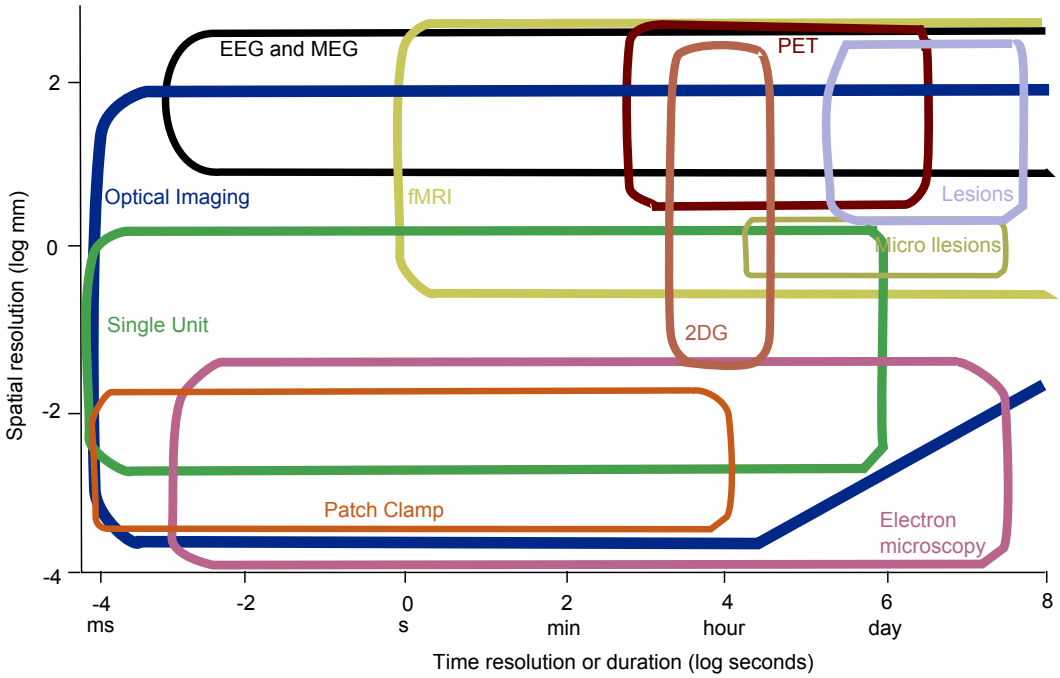
6.2 Temporal coding

Many studies, mostly in insects, have emphasized the importance of temporal network dynamics in encoding sensory memories (for recent reviews, see Gupta and Stopfer, 2011; Bathellier et al., 2010; see also section 7.1).

A prominent dynamic phenomenon in the OB are oscillations (Chapter 3 Wilson and Stevenson, 2006):

- gamma frequency (40–90 Hz): driven by mitral cell–granule cell circuits (c.f. Beshel et al., 2007),

Figure 6.1: ***Spatial and temporal characteristics of neuroimaging techniques.*** *Sides of rectangles indicate upper and lower limits of spatial and temporal characteristics of each technique. Adapted from Grinvald and Hildesheim (2004).*



- beta frequency (15–40 Hz): driven by OB – cortex feedback loops (c.f. Ravel et al., 2003), and
- theta frequency (2–15 Hz): driven by sensory input and respiratory cycle (c.f. Margrie and Schaefer, 2003).

Synchronization and phase–locking of M/T cell activity has been observed with the possible functional roles of increasing sensitivity, temporal binding, or multiplexing (Schaefer and Margrie, 2007; Schoppa, 2006; Friedrich et al., 2004; Schoppa and Westbrook, 2001; Kashiwadani et al., 1999, e.g.).

Many proponents of the temporal coding idea work on insect models, as opposed to researchers working on vertebrate models, who seem to put more emphasis on spatial coding schemes. This might as well have to do with spatial and temporal

dynamics of different methods of measurements at their disposal (c.f. section 6.1), but also could point to some differences in the insect and vertebrate systems.

The importance of time dynamics relative to spatial representations is not clear (c.f. Schaefer and Margrie, 2007; Khan and Sobel, 2004). According to the most radical epitome of the temporal coding idea, spatial patterns are largely epiphenomenal; on the other hand, some consider oscillations an epiphenomenon (c.f. Freeman, 1978; Fletcher et al., 2005; Singer, 1999). Most are more conciliatory, however, for example Knüsel et al. (2007) argued based on their model that temporal and spatial information could be complementary. Laurent (2002) discussed how bulbar dynamics on different timescales could serve functions in optimizing stimulus representation, such as sparsening, feature binding, and decorrelation, and how read–out mechanisms could make use of temporal patterns in different ways. However, he concluded that after optimization of representations, initial complexity could be reduced.

Rabinovich et al. (2008) argued that neural processing could be explained in terms of transient dynamics without any need for stable attractor states, a view that is contrary to the more traditional view of neural processing in terms of the attractor network model (c.f. Mozer, 2009; Lansner, 2009; Kesner et al., 2000; Hopfield, 1982). Niessing and Friedrich (2010) measured responses of 128 mitral cells in the zebrafish to two amino acids at different concentrations and found that while concentrations lead to small changes, switching between two odorants leads to abrupt network changes in the network states which can be accounted for by the attractor network model. Interestingly, this is what should be expected with categorical perception (Damper and Harnad, 2000).

Uchida and Mainen (2003) found that early information, as provided by the first sniff, is sufficient to reliably discriminate olfactory patterns. Spors and Grinvald (2002) described that spatial patterns at early and late stages of stimulation are largely maintained, however, that spatial patterns are refined over time and some glomeruli get lost in the process. This suggests that latency could play a role in the decoding of glomerular patterns (e.g. Zavada et al., 2011), however, it could mean as well that later changes to glomerular activities could be regulated by adaptation mechanisms and that steady–state information is enough for discrimination.

In the following, we will concentrate on spatial aspects of coding, where input olfactory stimuli patterns are classified into discrete patterns of activities. We will concentrate in this discussion on the glomerular level.

6.3 Population coding

Hinton and McClelland (1986) declared their belief in distributed representations and many people in the connectionist community have been repeating this credo. Experimental support for this argument is, however, not conclusive (for example Malach et al., 2002; Haxby et al., 2001)[1]. As for examples of distributed representations, movement directions in the monkey could be predicted by matrix multiplication of neuronal activities with their respective movement direction selectivity (Georgopoulos et al., 1986).

Haddad et al. (2010); Cleland et al. (2007); Youngentob et al. (2006); Lin et al. (2006); Wilson and Stevenson (2006); Rubin and Katz (1999) argued for population codes in the OB and there is evidence supporting their plausibility. Haddad et al. (2010) analyzed population activity of glomeruli and MT cells from different studies and concluded that the first principal component was highly correlated to approach or withdrawal, or pleasantness. They found it also correlated with many perceptual descriptors. Rubin and Katz (1999) showed that correlation between activity maps could explain differences in carbon chain length.

Johnson et al. (2005b, 2004, 2002, 1999); Rubin and Katz (1999) found correlations between differences of physico–chemical odorant properties and overall differences of glomerular response patterns. Johnson et al. (2004, 1999); Rubin and Katz (1999) demonstrated that differences in carbon number correlated with overall map differences, Johnson et al. (2005b) showed a significant correlation between distances of functional group position and differences between patterns.

In a scenario with population codes, each glomerulus would be broadly tuned across olfactory modalities and smell recognition would be the result of decoding the

[1] It should be noted that not finding topographic coding for a property is not a proof of distributed representation in that area. It could have to do with the stimulus or with the resolution of measurement.

combined activity of various glomeruli. Recognition can be achieved by integrating activations of a sufficiently broad subset of units. The population code for an odor quality would be the mean map over all activity maps corresponding to the quality (or the mean maps of relative levels of activation across the ensemble). The response profile of M/T cells is actually more narrow than that of glomeruli (Tan et al., 2010), a fact which rather speaks against such a broad integration.

Salinas and Abbott (1994) suggested that weighting the neurons would result in a more optimal code. If these weights become very unevenly distributed or sparse, we speak of a *spatial code* as opposed to a *population code*. If they are spatially clustered, we could speak of a *continuous spatial* or *local code* as opposed to distributed. Spatial coding refers to the case where encoding for a certain type of information is spatially restricted to a subset of the units of a population. Spatial codes could still be distributed, or spatially clustered. If spatial codes are clustered, we can speak of a labeled–lines coding (c.f. 6.4.1).

6.4 Spatial coding

Intensity and odor quality is coded by activity at the OB level. Odor identity is encoded by a unique spatial pattern across glomeruli. Odor concentration is related to the intensity and spread of this pattern. It is well established using different techniques that glomerular activations spread spatially out as a function of odorant concentration (Rubin and Katz, 1999; Cleland et al., 2007; Johnson and Leon, 2000a; Stewart et al., 1979). See fig. 6.2 for an illustration of glomerular activation patterns.

Many studies suggest a modular organization of the glomerular layer (c.f. Matsumoto et al., 2010; Johnson et al., 2005b,a; Takahashi et al., 2004; Johnson et al., 2004; Mori, 2003; Leon and Johnson, 2003; Johnson et al., 2002; Linster et al., 2001; Uchida et al., 2000; Johnson and Leon, 2000b,a; Johnson et al., 1999, 1998) in the sense of functional clusters of glomeruli, which respond to related features. Arguments for such an organization comes from a number of observations:

- glomeruli of related molecular specificity tend to be close on the glomerular layer,

- responses to odorants with common molecular features encompass overlapping regions,
- odorants activate sometimes different glomeruli within these regions.

Alonso and Chen (2009) define the term receptive field as a "portion of the sensory space that can elicit neuronal responses when stimulated." A prominent example for receptive fields are combinations of different features such as shapes and aspect ratios, which are detected by place fields in the hippocampus (Wills et al., 2005). In olfaction, receptive fields in terms of molecular properties have been called *molecular receptive ranges* (Mori and Shepherd, 1994). Please refer to Arzi and Sobel (2011); Murthy (2011); Khan et al. (2010) for recent reviews on the topic.

The properties of visual receptive fields can be explained from first principles (Lindeberg, 2010), however, similar characterizations of the olfactory system have not been done. As mentioned before in section 5.4.1, olfactory stimuli dimensions are still elusive (c.f. section 5.4; Sell, 2006; Haddad et al., 2008b) and there is no clear dimension along which to arrange a stimulus continuum. This missing metric for smell makes the study of odors particularly difficult and therefore also the characterization of receptive fields (Leon and Johnson, 2009; Haddad et al., 2008b).

General findings from studies on glomerular coding using different techniques could be summarized as the following (adapted from Khan et al., 2010):

- odorant molecules tend to elicits several foci of activations (Stewart et al., 1979),
- each of those foci consists of multiple glomeruli (Johnson et al., 1998, 1999),
- odorants with different chemical structure, shape, and perception generate distinctive patterns of glomerular activation,
- these patterns are bilaterally symmetric and very similar between individuals for a given odorant,
- increasing the ligand concentrations increases size of foci,
- glomeruli are molecular feature detectors (for example Galizia et al., 1999; Johnson and Leon, 2000a).

. Many groups did experiments on activation of glomeruli in response to a limited set of odorants which varied with respect to physico–chemical properties, which differed in several properties. The result from these studies is that varying some physico–chemical properties activated distinct clusters of glomeruli. Investigated properties which had particular influence on the localization of the activation peaks are functional groups (Johnson and Leon, 2000a; Johnson et al., 1998) and hydrocarbon structures (e.g. double–bonded, branched, cyclic, etc. Johnson and Leon, 2000b).

Many studies have shown using single mono–molecular odorants that odors can be discriminated based on distinct spatial responses at the glomerular level (e.g. Niessing and Friedrich, 2010). Simple odorant molecules tend to activate several regions in the OB (e.g. Spors and Grinvald, 2002) and each site of activation consists of multiple glomeruli (for example Johnson et al., 1995, 1998). In contrast, Lin et al. (2006) concluded that population responses to complex natural stimuli constituted roughly the sum of the responses to its individual components in a way that individual glomeruli signaled the presence of particular components. Johnson et al. (2010) found that glomerular activations of ecologically relevant odors were much more focal than those elicited by mono–molecular compounds.

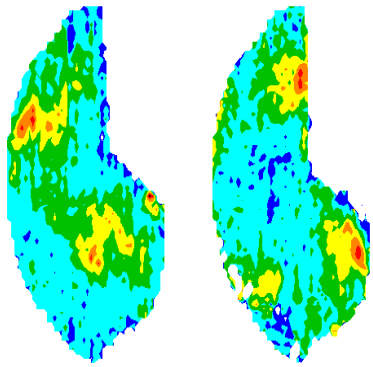
To give some more examples of spatial coding, Miklavc and Valentincic (2011) tested discrimination ability between 15 amino acids in the zebrafish and compared results to spatial overlap of most activated areas in glomerular response maps as per Friedrich and Korsching (1997). From their results, it can be argued that spatial activation patterns are the basis of discrimination ability. Auffarth et al. (paper IV; 2011b) demonstrated that physico–chemical properties can be reliably predicted from constricted glomerular zones. Auffarth et al. (paper V; 2011a) showed that qualitative odor perception can be predicted from small areas of glomerular activity patterns.

6.4.1 Topography

Continuous spatial codes can also be called *labeled–line coding*.

Topography is an isomorphic mapping between different spaces. As Thivierge and Marcus (2007) point out in their excellent review of topography in the brain, it

Figure 6.2: ***Glomerular activity maps.*** *Activity–related glomerular response to two odorants (left octanol, right propanol). Colors indicate (normalized) radioactive uptake. Odorants can be distinguished — to some extent — by the spatial activation pattern of glomeruli. Plot based on data obtained from Leon and Johnson (2006).*



is a common feature in the animal world, found in mammals as well as insects. It has been found at different levels of the brain that information is spatially embedded, in a way that the relative spatial structure of units is preserved between two regions (compare Thivierge and Marcus, 2007; Malach et al., 2002; Singer, 1994; Udin and Fawcett, 1988). Topography can be found in the visual (Wandell et al., 2005; Sereno et al., 2001; Tootell et al., 1988; Talbot and Marshall, 1941), auditory (Romani et al., 1982; Reale and Imig, 1980; Woolsey and Walzl, 1942), somatosensory (Marshall et al., 1941), and gustatory systems (Chen et al., 2011a). In particular, the spatial structure of representations in sensory cortices has been observed in many parts of the brain to locally reflect subspaces of stimuli that are behaviorally or perceptually relevant. Particularly well–known examples for this phenomenon come from studies of receptive fields in early visual and auditory processing, where afferent input from primary sensory neurons is spatially segregated (c.f. Humphries et al., 2010; Swindale, 2008). In vision, neurons are arranged across the cortical surface in an orderly orientation map of preferred stimulus orientations (called orientation pinwheels) (Ohki et al., 2006; Hubel and Wiesel, 1962). Topographic organization has been experimentally demonstrated in the primary auditory cortex (Merzenich, 1974), the retinotopic organization in the striate cortex (Tusa et al., 1978), the or-

ganization of the somatosensory system (Dawson and Killackey, 1987; Woolsey and Van der Loos, 1970). Prominent textbook examples for this principle are occular dominance columns, blobs in layers II and III in V1 of monkey, stripes in layers II and III of V2 of a monkey, and barrels in the primary somatic cortex (p. 207 Purves et al., 2004).

Topography is not tied to columns, since it is present even in areas of animals where there is no columnar organization (Kalatsky and Stryker, 2003). It is not limited to reflect spatial aspects of the input space, as illustrated in the case of the OB, the tonotopic map in the auditory cortex, or the specialization in the gustatory system. In taste, while the idea that regions of the tongue are exclusive for certain taste categories has been abandoned, it was found that there are fiber tracts responsible for specific gustatory sensations such as salty or sweet (Chandrashekar et al., 2006) and it has recently been found that a spatial map in the gustatory cortex encodes the five basic taste sensations (Chen et al., 2011a).

In olfaction, as for the glomeruli of the OB, Soucy et al. (2009) found a coarse topographic organization with respect to overall response profiles and many studies found molecular feature clusters (e.g. Matsumoto et al., 2010; Johnson and Leon, 2007; Mori et al., 2006a; Auffarth et al., 2011b; paper IV). Cleland and Sethupathy (2006) argued that the two–dimensional geometry of the OB could not accommodate the complexities of contrast–enhancement and made the argument that a projection to two dimensional space would result in a fragmented map. While it is not clear if the OB projection is two–dimensional (c.f. fig. 5.4), topography is complemented by splits, disproportionate magnifications, and other transformations (after Thivierge and Marcus, 2007, ; Kaas, 1997; Sereno, 1991). Inhibitory long–range connections could be part of these constraints that establish symmetry (c.f. section 5.3.5 for symmetry in the olfactory system and Kaschube et al., 2010). As for the OB, Johnson and Leon (2007) observed that the spatial progression of some properties on the glomerular layer is often disrupted by unresponsive glomeruli. They provided two possible reasons for that: 1. some related odorants were not tested 2. unrelated specificity might be present in these (intermittent) domains.

In the piriform cortex, a topographic order has not been found, however, it was found in projections to the amygdala and the anterior olfactory nucleus (compare

section 5.3.6).

Interestingly, the formation of topographic maps can be explained by activity–dependent principles, compare section 5.3.3.

It has been argued that such an organization helps to minimize wiring length (Chen et al., 2006; Buzsáki et al., 2004). The functional implications of topographic organization are not clear, however, (e.g. Kaas, 1997). Different arguments have been made in favor of a computational advantage conveyed by this principle, which fall into two categories (after Thivierge and Marcus, 2007): (i) local computation (Sporns, 2010; Hilgetag et al., 2000) and (ii) faithful representation of a 1–1 relationship (Pulvermüller, 2005; Simmons and Barsalou, 2003; Friston, 2002; Barlow, 1981).Further, such an organization could imply a sparse code, where few units encode particular patterns. Such a sparse code could in itself confer several advantages (c.f. Olshausen and Field, 2004).

It was suggested that inhibition between glomeruli decorrelates M/T activity and sharpens mitral cell response profiles and that this could be supported by a topographic organization (Aungst et al., 2003; Friedrich and Laurent, 2001; Mori and Shepherd, 1994). Niessing and Friedrich (2010) studied M/T activities in the zebrafish in response to amino acids (natural food odors for zebrafish) for which it is known that they have highly similar glomerular patterns and found differential cell activity. They argued that chemotopic coding could support decorrelation of similar odors.

6.4.2 Spatial progressions

Johnson et al. (1999, 2004); Ho et al. (2006) remark that increasing the number of carbon atoms, glomerular activations shift gradually to ventral locations (*chemical progression*; Johnson et al., 1999, 2004). This shift occurs mostly within regions responsive to functional groups, however, is also present in odorants without functional groups (Ho et al., 2006). We found such a progression in a systematic study based on 362 odorants (paper IV; Auffarth et al., 2011b).

At the same time, analysis of carboxylic acids with different hydrocarbon structures demonstrated a ventral progression medially with molecular length (as opposed to hydrophobicity or volume; Johnson and Leon, 2000b). Johnson et al.

(2004) suggested chemotopic progression as an organizational principal in olfactory processing. Interestingly, in the epithelium, more hydrophilic, less volatile, heavy odorants absorb more readily into the olfactory mucosa and would therefore project more dorsally, while more hydrophobic and shorter or lighter molecules could absorb more posterior and thereby project to peripheral and ventral zones (Schoenfeld and Knott, 2004).

Similarly to what Johnson et al. (1999, 2004) observed about functional groups and carbon number, Yoshida and Mori (2007) proposed primary and secondary odorant categories, where secondary categories could serve to enhance category–profile selectivity. As primary properties they suggested sulfides, alcohols, methoxypyrazines, 6–carbon and 9–carbon green–odor compounds, aldehydes, ketones, isothiocyanates, terpene hydrocarbons, esters, terpene alcohols, alkylamines, acids, lactones, and phenol and its derivatives. Our results align well with these suggestions (paper IV; Auffarth et al., 2011b).

It is further known that carbon–chain length is related to discrimination ability (Cleland et al., 2002) and there is evidence for a progression by carbon–chain length (paper IV; Auffarth et al., 2011b; Johnson et al., 2004). This indicates that spatial segregation (spatial distance between representations) could have behavioral correspondences.

6.4.3 Principles that determine the localization of representations

A particular distinction of the olfactory system is the relative lack of intrinsic spatial topology of the stimulus as compared to other senses (or at least it is not understood). While visual and tactile stimuli have an spatial dimension to them, auditory stimuli get converted to a spatial domain by frequencies.

Generally, factors that determine the localization of representations in the OB are the following (compare Mori and Sakano, 2011):

- molecular features,
  - chemotopy, or (related)
  - odotopy.
- rhinotopy,

- perceptual relevance (c.f. section 6.5), and
- genetic identity of axon guidance cone (c.f. section 5.3.3).

There is accumulated evidence for continuous spatial zones responsive for certain groups of odorants based on data from different techniques. It is well established that some molecular odorant properties map differentially to clustered spatial locations (for example Matsumoto et al., 2010; Mori et al., 2009; Johnson and Leon, 2007; Mori et al., 2006b; Lodovichi et al., 2003; Meister and Bonhoeffer, 2001; Vassar et al., 1994) and our results from a systematic large–scale study of glomerular activity indicated that coding of some molecular properties is organized in continuous zones and locally very restricted (paper IV; Auffarth et al., 2011b). This map–like feature at the level of the OB that seemingly represents molecules as perceptual objects (Haberly, 2001; Mori and Shepherd, 1994) has led to comparisons of the OB to sensory cortices (Schoenfeld and Knott, 2004; Haberly, 2001; Johnson and Leon, 2000c).

Furthermore, the relative spatial locations between representational areas corresponding to odorant categories could be related to chemical similarity (Johnson et al., 2004; Cleland et al., 2002; Uchida et al., 2000). This principle has been referred to as *chemotopy* or, *spatial progression*, in the case of a spatial shift related to molecular properties. A related concept is *odotopy*, which refers to a spatial arrangement by odotopes.

In the olfactory system, there is a spatial order of glomeruli related to the distribution of ORNs at the epithelium which is referred to as rhinotopy (Johnson and Leon, 2007; Miyamichi et al., 2005; Schoenfeld and Knott, 2004; Astic and Saucier, 1986).

6.5 Representation of perceptual dimensions

Where physico–chemical properties are mapped to perceptual qualities is unclear. A perceptual encoding of representations has been found previously in the epithelium (Lapid et al., 2011) and the piriform cortex (Howard et al., 2009).

Youngentob et al. (2006); Johnson and Leon (2000d) argued that qualitative odor perception is, in fact, determined by glomerular activity patterns. They described

that different glomerular activity patterns, elicited e.g. by increased concentration can lead to qualitatively different odor percepts. Haddad et al. (2010) showed that the first principal component of population activity of glomeruli and MT cells was highly correlated to approach or withdrawal in animals, or pleasantness in humans. Cleland et al. (2002); Ho et al. (2006); Linster et al. (2001, 2002) found correlations between behavioral measures and pattern similarity. We found that perceptual categories are mapped to continuous spatial zones (paper V; Auffarth et al., 2011a),

Studies in rats and mice have shown that different types of behavior, e.g. defensive behavior towards predators, aversion or attraction towards food, can be related to the chemical categories of odorants and generally supports the behavioral relevance of molecular feature combinations and glomerular domains (paper VI; Auffarth et al., 2011c; Sakano, 2010; Raman and Gutierrez-Osuna, 2009; Kobayakawa et al., 2007; van der Goes van Naters and Carlson, 2007; Stockinger et al., 2005; Dielenberg and McGregor, 2001). A study by Kobayakawa et al. (2007) suggested that the OB of mice consists of at least two different functional modules, one for innate odor responses and one for (associatively) learned odor responses. Similar spatial behavioral organization is also known to occur in insects (for example Semmelhack and Wang, 2009). Changing locations of glomeruli in mice can result in behavioral impairments, in spite of persistent physiological activations (Adam and Mizrahi, 2010).

Chapter 7

Data and methods

In this chapter, we will introduce some methods and data sets that were used throughout.

7.1 Glomerular Response Archive

The glomerular response dataset was used in papers V and VI. In these studies, we used a set of 2–deoxyglucose (2DG) autoradiography images of glomeruli covering the entire lamina of the rat olfactory bulb (Leon and Johnson, 2006; Johnson et al., 2006, 1999). These response images are part of a database collected by Michael Leon's group over the course of many years. They have formed the basis of many published studies and they are available for download from their public database. See fig. 6.2 for an illustration.

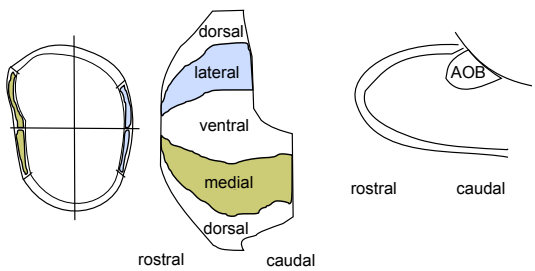
2DG is taken up by the glucose transporters of cells and uptake correlates with cell glucose matabolism and hence activity in neurons. Generally in 2DG measurements, the 2DG is labeled by a radioactive substance such as tritium or carbon–14 which can then be measured in autoradiography.

In the experiments in Leon's lab, they stimulated unanesthetized and freely respiring animals with odorants. The animals were then dissected and imaged. Each of these images corresponds to glomerular responses to one particular odorant. Leon and Johnson took averages over the left and the right bulbs and averaged over arrays from all animals exposed to the same concentration of the same odorant (roughly

around 3 to 5 animals for each combination). It is known that patterns are very similar across animals and that glomeruli are at stereotypic positions (Soucy et al., 2009; Fantana, 2006). According to Johnson et al. (2006), the 2DG method is capable of resolution down to a single glomerulus, however, averaging over multiple animals results in a loss of apparent spatial resolution. The 2DG method provides a single (averaged) response map for each odorant presentation and is therefore unable to resolve the temporal dynamics of activity. According to Rubin and Katz (1999); Guthrie and Gall (1995); Slotnick et al. (1989) glomerular activations are relatively constant over time, indicating the absence of temporal dynamics might not be a limitation, however, over time, spatial patterns become refined and some glomeruli are not activated anymore (Spors and Grinvald, 2002), and Wilson and Stevenson (2006, ch. 5) cautioned that after initial sampling, subsequent changes might reflect an adaptation process. For a discussion of differences in method to other studies and their implications, please compare Johnson and Leon (2007).

The ventral–centered chart of the glomerular layer in fig. 7.1 denominates locations in anatomical terms of the rat olfactory bulb.

Figure 7.1: ***View types of glomerular images.*** *This figure visualizes locations in anatomical terms for the glomerular layer in dorsal and ventral–centered charts of the rat olfactory bulb. These charts have been described as "rolled-out maps" (Leon and Johnson, 2006). They open dorsally and are rotated so that dorsal is located at the top and bottom boundaries of the charts, lateral is located in the upper half, medial is in the lower half. Rostral is to the left and caudal to the right. This is adapted from Johnson et al. (1999).*



Nawroth et al. (2007) analyzed the energy metabolism within GLs of the OB and concluded that, while presynaptic inhibition of the sensory axon terminals by periglomerular cells modulates activity ranges, most of the energy budget is related

to axodendritic transmission between axons of primary and dendrites of secondary neurons. It should be cautioned, however, about inferring M/T cell firing patterns from these data. There are papers that show that there is a high correlation between glomerular imaging signals to M/T cell activations (e.g. Moreaux and Laurent, 2008) and further, results from tracing studies indicate a functional compartmentalization by receptor channels throughout glomerular and M/T layers (Willhite et al., 2006). However, M/T cells receive excitatory and inhibitory input from other glomeruli (Najac et al., 2011; Kim et al., 2011, ; c.f. section 5.3.4). Therefore, we do not pretend to know about the spatial patterning of responses on the M/T layer.

7.1.1 Molecular descriptors

For each of the activation maps we had information about which odorant they correspond to and some additional descriptive information, which was also provided by the Leon Lab. These descriptors, about 200 in total, include physico–chemical odorant properties, as well as perceptual properties associated with the odor. Properties are of continuous and binary type. Continuous properties include molecular length, height, and weight. To give some examples of binary properties, these concerned cyclization (whether an odorant is alicyclic, aromatic, polycyclic, or heterocyclic, respectively), bond saturation (whether an odorant is alkene, alkane, or alkyne), and functional groups (whether an odorant is ester or lactone, amine, carboxylic acid, contains sulfur, contains halogen, is a ketone, alcohol, or phenol). Perceptual properties are all binary and include flavors such as sweet, camphoraceous, floral, and minty.

For the classification study reported in paper IV, we took into account 13 binary molecular properties where at least four images were available. These were sulfur–containing compound, alkyne, alkane, alkene, amine, carboxylic acid, aromatic, ketone, polycyclic, ester+lactone (being both an ester and a lactone), alcohol+phenol (being both an alcohol and a phenol), heterocyclic, and alicyclic.

We defined one additional property as elongation, which we computed as the variance over molecular length, surface area, and molecular depth.

7.1.2 Data pre–processing

In our pre–processing of activation maps, we started with 472 maps, of which some represented responses to identical odorants in different concentrations. By visual inspection, we eliminated five highly saturated images where most or all of the glomeruli were activated and took means over the responses corresponding to the same odorant in different concentrations (222 maps in on average 3 concentrations). Ten maps had to be discarded because of missing odorant information. This selection left us with 308 point maps, each of which corresponded to odor–induced glomerular activities. Pixels with missing values, caused by loss of tissue during cryosectioning, others due to loss of tissue during removal of the bulbs from the skull using microdissecting scissors, were removed, which left us with 1834 pixels per map. We mean–centered all pixels and normalized deviations to standard unit to compensate for differences in absolute pixel intensities.

For some binary properties, there were many associated activity maps, for others very few. We only took into account binary properties where the less frequent case was represented by at least four maps.

7.2 Odorant perceptual properties

In papers V and VI, we analyzed representations according to perceptual categories of odorants. In paper V, we took a set of odorant perceptual properties (sometimes called *organoleptic properties*) provided by Leon and Johnson (2006) together with the odorants.

For paper VI, we extracted perceptual odorant descriptors from flavornet (Acree and Arn, 1998)[1], a public internet resource about basic volatile compounds that humans experience in their environment. Examples for these odor descriptors are sweet, camphoraceous, floral, or minty.

In both papers, we grouped organoleptic properties into odor categories, which we will henceforth call *odor qualities*. These are odor perceptual categories, as described by Zarzo (2008). These categories are florals, cleaner (cleaning agent),

[1] Available at `http://www.flavornet.org/`

foul, woody, medicinal, nutty/spicy, balsamic, fruity, alcohol, oily, herbaceous, musk, vegetable, and green.

7.3 Localization of coding zones

To determine the activation loci for certain dimensions of the odor stimuli, such as odor qualities or physico–chemical properties, we applied a statistical test. The idea behind this test was to check, whether activations at a pixel increased significantly when the odorant exhibits a particular property with respect to when it does not.

For each molecular property, we tested whether a pixel showed significant differences with respect to the property. For binary properties (such as whether it was an aromatic chemical or not, or sweet smell or not) we compared activations for stimuli where the binary property was given, A_P, against response patterns where the binary property was not given, $A_{\neg P}$. This was done using the *Wilcoxon rank–sum test* (also called the Mann–Whitney U test), a non–parametric statistical test, which assesses whether two samples come from the same distribution (null hypothesis).

For continuous properties the procedure was more involved, such as described in the following. We discretized properties by grouping their values into bins, taking bin numbers k as first guess from Sturges' formula (compare Wand, 1997), which is as follows:

$$k = \lceil \log_2 n + 1 \rceil, \tag{7.1}$$

where braces indicate the ceiling function, n are the number of observations (values in the vector).

Then k was manually adjusted such that in each bin there were at least roughly 5% of activation maps.

We could have used different methods, such as histogram equalization, density models, or optimization of k according to some measure, however decided, for simplicity, to stay with this method.

We then applied the procedure with boostrap and Wilcoxon rank–sum test for differences between activations in response to property values in a particular bin versus activations in response to values out of the bin, i.e. testing whether points corresponded to different ranges of the distribution of the chemical property.

7.3.1 Wilcoxon rank–sum test

The Wilcoxon rank–sum test includes the calculation of a statistic U. The distribution of U is known for the null hypothesis, which assumes that the two distributions are not different. For a detailed explanation, refer to Ott and Longnecker (p. 305 2008):

The test compares two data samples, vectors S_1 and S_2, which are of equal length N. A general way to compute the Wilcoxon rank–sum test is to obtain a value, the z–value, which can be looked up in statistical tables. This can be done as follows:

1 We start by calculating a vector of differences between entries in the two samples: $Z^i = S_1^i - S_2^i$.

2 Then we compute a rank vector R of these differences by sorting absolute values of Z in descending order.

3 A value W, with weighted entries of Z is given by this formula:

$$W = \sum_{i=1}^{n} -1^{1-\phi_i} R_i \tag{7.2}$$

where ϕ is an indicator function:

$$\phi_i = I(Z_i > 0). \tag{7.3}$$

4 This gives the z–value, which can be looked up in statistical tables, according to this formula:

$$z = (W \pm 0.5)/\sqrt{N(N-1)(2N+1)/6}. \tag{7.4}$$

Using the Wilcoxon ranked–sum test is analogous to applying the student t–test on the data after ranking over the combined samples. It has the advantage over the t–test of not assuming normality.

7.3.2 Bootstrapping

For some properties, we had only very few maps that corresponded to them. To account for statistical variations in these distributions and differently sized vectors

we applied a bootstrap (Efron, 1982) resampling procedure for all tests. The purpose was to estimate p–values of the statistical test.

In bootstrapping, a statistical analysis is repeatedly applied to subpopulations of the same size, generated by sampling from the original population with replacement. Bootstrap methods can be used for hypothesis tests and for regression analysis and allow the estimation of distributions of almost any statistic where only few samples are available. For a detailed explanation, refer to Chernick (2011).

The combination with bootstrapping gives our method more robustness to outliers and allows the two samples to be of arbitrary (unequal) sizes.

7.3.3 Definition of representation

For each pixel, we will define *receptive fields* as the range of stimuli for which we can show they evoke excitatory response at significance level.

Definition We say that points are coding for a (binary) property if the null hypothesis could be rejected at the 5% significance level. More formally, the coding of a property can be expressed as:

$$\text{coding} := (\hat{p} \leq 0.05) \times (1 - \hat{p}) \times (-1)^{2^{(\text{bigger}>0.5)}} \tag{7.5}$$

where $\hat{p}$ stands for the bootstrapped p–value of the statistical test. *bigger* takes 1 if the bootstrap estimated of the distribution means $\hat{\mu}_{A_P} > \hat{\mu}_{A_{\neg P}}$ and 0.

The obtained value indicates how strong odor–induced activity is associated with a property (or its absence for negative values). If statistical significance is below a threshold, the value is set to 0. We only took into account values above 0.

7.3.4 Size of coding zones

In order to determine the size of coding area for a property, we took the number of points that responded significantly different when a property was given as when the property was absent. Skewed distributions for some properties could have an impact on how many points are found to be significantly related to a property. Therefore it is important to note that for the properties under consideration, data

availability (number of images corresponding to presented molecular properties) and size of coding zone showed no significant Pearson correlation ($\rho = 0.33$, $p = 0.27$). We took into account 13 binary molecular properties, where at least four images were available.

7.4 Relevance of physico–chemical properties

In paper IV, support–vector machine (SVM) classification was utilized as a non-probabilistic binary linear classifier to test the relevance of physico–chemical features. A SVM was chosen because its superior performance compared to other classification techniques (e.g. Kotsiantis and Zaharakis, 2007) makes it an algorithm of reference. We take the classification performance as a quantitative measure of the structure–activation relationship between activations of glomeruli and odorant features. We take its result as indication of relevance of molecular properties to glomerular coding.

The scarceness of data in the classification problem for some properties lead to problems with the classifier. We implemented a MATLAB wrapper for SVM–light (Joachims, 1999) and tried libSVM Chang and Lin (2001), which are likely the most popular SVM implementations, however they failed to give consistent results over repetitions. In the end, we settled on a linear SVM (SVM Cortes and Vapnik, 1995), in a MATLAB implementation by Miquel Tarzan, a lab colleague in Barcelona. We found this implementation to be more robust than other SVM implementations, possibly because of the use of Mathwork's implementation of quadratic programming (quadprog function).

Classification was performed on a 32–core linux cluster, a beowulf, that I had assembled and installed (Auffarth, 2009). For classification, glomerular activations were taken as input vector and each binary property (present vs. not present) as binary target. In each of 10 iterations we randomly sampled half of the activation maps as training set and took the other half as test. We distinguished between two experimental conditions: 1. best points — classification using most representative points, and 2. random baseline — classification using randomly sampled points.

We classified molecular properties by glomerular activations as input in order

to estimate their impact on early olfactory coding (relevance). The logic behind is that properties that greatly change activations at the OB level should be easier to classify. Knowing the relevance of molecular properties could provide insight into early coding of chemical information and provide vital clues for discerning which properties are functional in determining the degree of interaction between an OR and odorant molecules. We define relevance as the best classification performance from either most representative points or random baseline whichever was higher.

For the first experimental condition, for each property, we sorted points in descending order by their relevance with respect to the property (p–values from Wilcoxon rank–sum test) and then classified taking the best n points, with $n \in N = \{1, 5, \ldots 30, 45, 50, 60, \ldots 150, 200, 300, \ldots 1700, 1834\}$ (36 steps).

As a random baseline, for each property, we took the same intervals N, but randomly sampled points. We averaged over 250 random subsamples of points for each interval.

We computed the area under the ROC curve (AUC; compare Bradley, 1997) as performance criterion, which measures the fraction of true positives against the fraction of false positives. It has the advantage to be unbiased by skewed class distributions, which are a particular problem in our data set.

7.5 Perceptual distances

We investigated how perceptual categories are represented in the olfactory bulb responses in paper VI. We made a distinction between *spatial coding*, the situation, where specialized local encoders exist for certain information, and, on the oher hand, *population coding*, where information is broadly distributed over the responses of the population. We analyzed and compared glomerular responses to human perceptual categories.

Zarzo and colleagues (Zarzo, 2008; Zarzo and Stanton, 2009) published analyses of two studies, perfumers' odor perception space (BH; Boelens and Haring, 1981) and cross–cultural odor similarity ratings (Chrea, 2004). In order to know, how well the coding as identified by the coding centers reflected perceptual orderings as reported in the literature, we mined PCA plots in the two papers by Zarzo,

which indicated perceptual distances in the first two principal components between odorant qualities. This provided us with a pair–wise distance matrix between two sets of perceptual odor categories.

We extracted two distance matrices which came from the two mentioned datasets, D_{Chrea} and $D_{\text{BH small}}$. *Chrea* corresponded to odor categories floral/cosmetic, cleaner, foul/musty, woody, medicinal, nutty/spicy, balsamic, and fruity. *BH small* corresponded to categories floral, woody, medicinal, balsamic, fruity. We compared matches to these perceptual spaces from both population activities of the entire glomerular layer and from spatial codes in order to see which reflected better these perceptual orderings.

In order to obtain codes for population responses, we took the mean map over all activity maps corresponding to the same odor quality (compare Cleland et al., 2007; Rubin and Katz, 1999; Lin et al., 2006).

Thus, a population code for a given odor category A can be written as $\vec{v}_A = (\langle x_1 \rangle_A, \ldots, \langle x_{\text{nglom}} \rangle_A)$ where A is the set of odorants representing category A and $\langle x_i \rangle$ stands for the mean response of glomerulus i averaged over all odorants belonging to A. As an ordering between properties we calculated the Euclidean distances between these mean–maps, thus obtaining pairwise distances between odor qualities based on population code, D_P.

Johnson and Leon (2000a); Johnson et al. (1998, 1999) used various dissimilarity indices and Johnson et al. (2002, 2004, 2005b) used the Pearson correlation coefficients between glomerular response patterns.

As for the ordering between spatial zones (as per explained definition), we applied the Hausdorff distance (compare Alt et al., 2003), which measures distances between two–dimensional shapes and therefore incorporates coding center distance (similar to Euclidean distances), but additionally information of shape, size, and orientation match. We applied the modified Hausdorff distance function (Dubuisson and Jain, 1994) between vertices of pairs of encoding zones. Vertices consisted of points that were found to be responsive to odor categories.

Informally, the Hausdorff distance is the farthest distance of closest points between two sets. Formally, given X and Y, two non–empty subsets of a metric space

(M, d), their Hausdorff distance $d_H(X, Y)$ is defined as follows:

$$d_{\mathrm{H}}(X,Y) = \max\{ \sup_{x\in X} \inf_{y\in Y} d(x,y), \sup_{y\in Y} \inf_{x\in X} d(x,y) \}, \tag{7.6}$$

with sup and inf representing the supremum and infimum, respectively.

Thus, we obtained a matrix of pairwise differences between properties based on coding maps, D_S.

We normalized each of the three matrices of pairwise distances D_P, D_S to unit sum and computed the sum of the absolute error between both of them and D_{Chrea} and $D_{\mathrm{BH\ small}}$.

We also added spatial and population information linearly with the same weight to see if combined they provided a better fit to the perceptual space. For the baseline, 100,000 sets of points were sampled from random uniform distributions. Then the distances from their pair–wise distances to perceptual space was computed.

7.6 Model of the early olfactory system

In paper VI we presented a model of the early vertebrate olfactory system consisting of ORs, ORNs, and GLs.

7.6.1 Receptors and receptor neurons

As commented before, there is no simple scale or known dimensionality to olfactory perception. Therefore one solution to order odorants is representing them by a large number of molecular descriptors. Haddad et al. (2008a) presented a set of 32 physico–chemical descriptors, derived from an initial set of $1,664$ descriptors, that were shown to reflect the variability of the bulb and antennal lobe population responses. In simple terms, this means that odors that cause similar responses are proximal in this space and odors that elicit dissimilar responses are distant. They supplied a dataset of 447 odorants described by these 32 properties as supplementary material with their paper, which we used in this study.

We devised a set of synthetic (or *virtual*) receptors. These odorant receptors should be distributed to capture the variance of the physical space and each be

placed to recognize biologically relevant regions (compare Schmuker and Schneider, 2007; Sánchez-Montañés and Pearce, 2002, for similar concepts). We applied the fuzzy c–means algorithm (Bezdek, 1981) to draw cluster centers at locations in the 32–dimensional space. In this way, each OR responds to ligands that occupy a neighborhood in physical space as described by molecular descriptors. The closer the combination of molecular properties of an odorant to the center of the receptive field of the receptor, the higher the response. Each OR can be described by its center in the 32–dimensional space and its affinities to odorants based on the distance relation in the 32–dimensional space. OR–odorant affinity relationships are indicated by the circle radii in fig. 2, paper VI.

We modeled ORN responses to ligands at a given concentration after Sandström et al. (2009) as a a sigmoidal function of the OR–ligand affinities. The response $R_i(C)$ of ORN i to a ligand at concentration C is expressed as the product of a term A_i that represents the amplitude and a term that includes the ligand responses of ORs. a_i is the affinity of receptor to the ligand and h is the gain (steepness) of the ORN response curve.

$$R_i(C) = A_i \left(1 - \frac{1}{e^{h_i a_i C}}\right) \tag{7.7}$$

The response of ORN population i is expressed as the product of a term μ_A and a term that includes the ligand responses of homogeneous ORs. The mean frequency of responses was taken to be as in this formula (adapted from Sandström et al., 2009):

$$\mu_{R_i}(C) = \mu_A \left(1 - \frac{1}{e^{\mu_h a_i C}}\right) \tag{7.8}$$

We set the mean amplitude, μ_A, of ORNs to 1 and mean gain μ_h to 1.4 (after Sandström et al., 2009). Please see fig. 3, paper VI, for an illustration of ORN dose–response curves.

7.7 Dimensionality reduction

We used multi–dimensional scaling (MDS) (Kruskal, 1964) in papers V and VI. A recent review and comparison about dimensionality reduction, including MDS,

can be found in Bunte et al. (2011). Generally dimensionality reduction is about reducing a dataset from $\mathbb{R}^N \ni x_i \to y_i \in \mathbb{R}^k$, where k is often chosen as 2 or 3. One difference between dimensionality reduction methods are the definition of similarity and employed optimized techniques. Consequently, these techniques also differ with respect to the preservation of distances, however, there is no universally accepted definition of topology–preservation (c.f. ch. 5 Lee and Verleysen, 2007).

Multidimensional scaling makes few assumptions about the structure of data and preserves the distance relationships among data samples. In MDS, the topology is preserved by minimizing (e.g. using gradient descent) the (Euclidean) pair–wise distances in the original data set $d_\chi(x_i, x_j)$ and pair–wise distances in the projection $d_\varepsilon(y_i, y_j)$:

$$E_{\mathrm{MDS}} = \sum_{ij} w_{ij} \left(d_\chi(x_i, x_j) - d_\varepsilon(y_i, y_j)\right)^2. \tag{7.9}$$

The weights w_{ij} can be set to $w_{ij} = 1/d_\chi(x_i, x_j)$ (Lee and Verleysen, 2007).

7.8 Axon growth

We clustered ORN axon projections by a biomimetic method described in Lansner et al. (2009). Using multidimensional scaling (MDS) (c.f. section 7.7), axons can be put in a lower–dimensional space where their locations are defined by distance relations based on co–activation. In this way, the distances between glomeruli reflects regularities in the physical odor space. For distance relationships, we computed Pearson correlations between vectors of response activities of ORN populations to all odorants.

We reduced the resulting matrix by MDS to three dimensions and obtained coordinate points corresponding to each olfactory axon bundle. In total, this operation is similar in principle to the self–organizing map used earlier as ORN convergence model (Gutierrez-Osuna, 2002b; Schmuker and Schneider, 2007) and is consistent with the chemoaffinity hypothesis (Sperry, 1963).

Chapter 8

Results about characterization of the olfactory bulb

As discussed before in chapter 5, glomeruli in the olfactory bulb (OB) are first processing stations, where odor representations are noise–filtered, sharpened, and possibly re–organized. Glomeruli are also gateways to secondary neurons and to higher–order representations. Therefore, learning about glomerular representations can possibly give important insights into olfactory processing. In this chapter, results from papers IV, V, and VI are summarized. These papers mainly deal with the spatial organization of olfactory information at the glomerular level.

The results presented in this chapter will then be discussed in chapter 9.

8.1 Introduction

The relationship between molecular properties of odorants and neural activities is arguably one of the most important issues in olfaction and the rules governing this relationship are not clear. In paper IV, we investigated glomerular representations of odorant properties. We compared the relevance of these properties and the size of their coding zones. This was done on a dataset of nearly 400 odorants which covered the whole array of glomeruli (c.f. section 7.1). In chapter 7, the methods used for data analysis are described, including a nonparametric statistical test and

a support-vector machine classification study.

As discussed in chapter 5, representations of perceptual qualities have previously been found in the piriform cortex, however, several recent studies indicate that the olfactory bulb code reflects behaviorally relevant dimensions spatially as well as at the population level. For paper V, we applied the same statistical analysis techniques on the odor response images as in paper IV. The purpose in that study was to understand odor quality coding in the brain. We used several techniques in the analyses to compare hypotheses of population and spatial coding (c.f. section 7.5). The most important and novel result reported in paper V was that we found representational areas for several perceptual qualities.

In paper VI, we modeled the early olfactory system, specifically the levels from the olfactory epithelium to the olfactory bulb. We derived a glomerular organization based on a set of real–world, biologically–relevant stimuli. Axons were ordered based on correlations of sensory neurons and thereby we found a spatial organization of glomeruli, which we analyzed with similar techniques as in paper V. Our main result was to show that this simple spatial organizational principle of glomeruli plausibly can explain a range of phenomena, including perceptual ordering comparable to that shown in paper V.

The rest of this chapter is ordered by papers.

8.2 Results on physico–chemical representations in the rat glomerular map

8.2.1 Localization of coding zones

In figures 8.1 and 8.3, localizations of coding areas for several physico–chemical properties are displayed. These results were obtained by applying the definition for representation and the corresponding statistical test described in section 7.3. These properties included molecular bonds, cyclization, functional groups, molecular length, and carbon number. In figure 8.3, spatial progressions with molecular length and carbon number are shown. These specializations to combinations of physico–chemical properties show a local clustering of receptive glomeruli. As can be seen in figure 8.3, the spatial representations for the physico–chemical properties are

Figure 8.1: ***Representations of physico–chemical properties****: Maps with coding zones for various odorant properties. Loci of the* 13 *binary properties (described in section 7.1), grouped into basic dimensions of molecular bonds, cyclization, and functional groups. The colors in the figures indicate zones that code for a specific combination of different binary properties as determined by the statistical method described in section 7.3. The numbers in the legends are explained in this caption.* ***a*** *Example map showing loci coding for the alkane property. 1 marks the coding zone for alkane.* ***b*** *— A map of odorant bond properties. The numbers stand for: 1 alkane 2 alkene 3 alkane, alkene 4 alkyne 5 alkane, alkyne 6 alkene, alkyne 7 alkane, alkene, alkyne.* ***c*** *— Map of two cyclization properties. The number code is as follows: 1 aromatic 2 alicyclic 3 aromatic and alicyclic.* ***d*** *Areas for two more cyclization properties. The number code is as follows: 1 polycyclic 2 heterocyclic 3 polycyclic and heterocyclic.* ***e*** *Coding zones for three functional group properties. The numbers stand for: 1 amine 2 ketone 3 amine and ketone 4 alcohol–phenol 5 amine and alcohol–phenol 6 ketone and alcohol–phenol.* ***f*** *Loci for the three other functional group properties. The numbers: 1 ester+lactone 2 carboxylic acid 3 ester+lactone and carboxylic acid 4 sulfur-containing compound 5 ester+lactone and sulfur-containing compound 6 carboxylic acid and sulfur-containing compound 7 ester+lactone, carboxylic acid and sulfur-containing compound.*

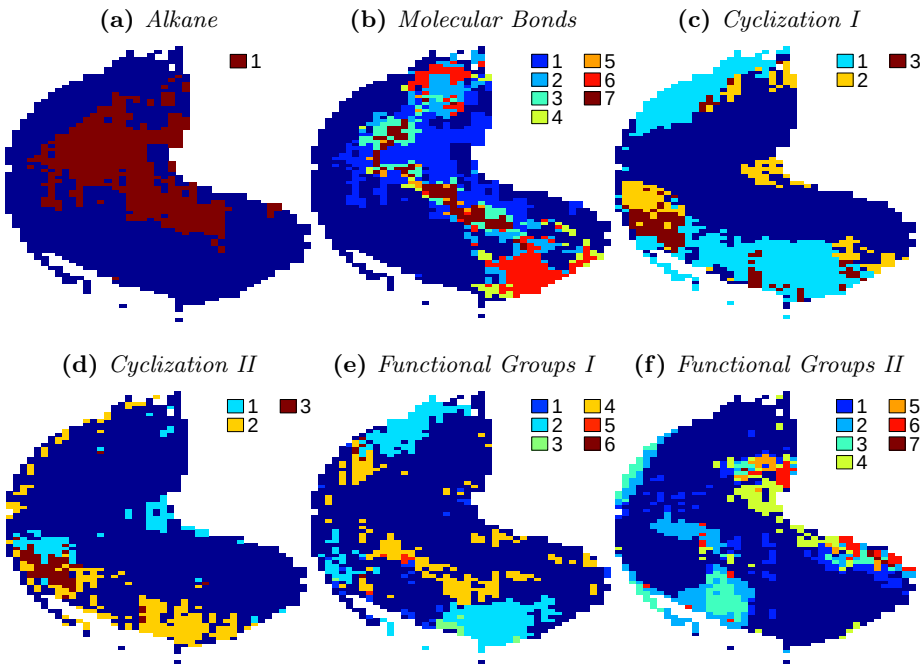


Figure 8.2: ***Representations of physico–chemical properties: a — surface area, b — elongation, c — water solubility.*** *Color explanation: yellow in figure a indicates a large odorant surface area, orange indicates a small surface area. Orange in figure b indicates molecules that are elongated (c.f. section 7.1, yellow indicates coding for more roundish molecules. In figure c, red indicates high water solubility, yellow indicates low solubility.*

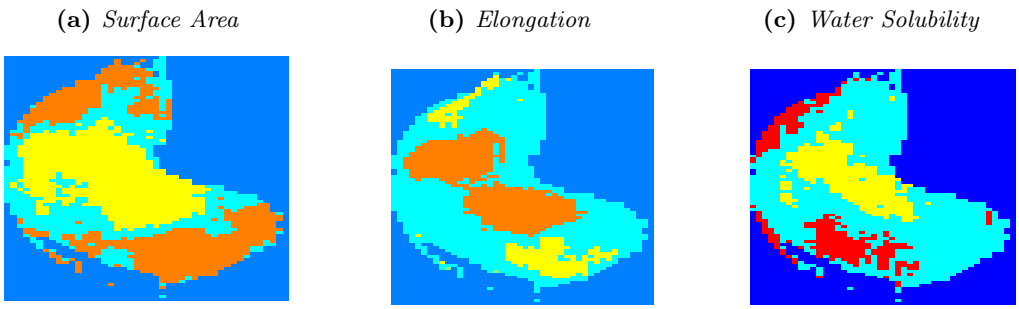
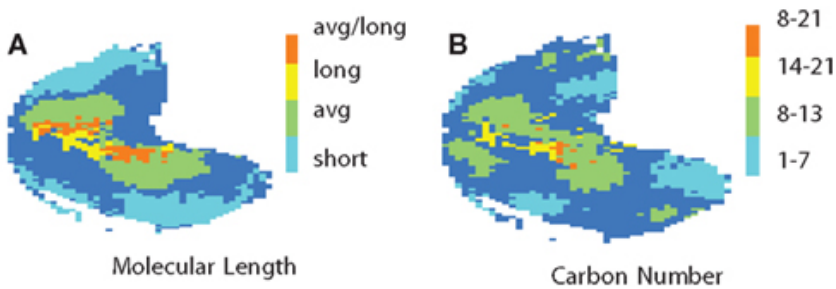


Figure 8.3: ***Representations for physico–chemical properties.*** *In* ***A*** *Molecular length across the bulb. See the legend for the color code.* ***B*** *Representations of carbon numbers. The legend explains the color code.*



organized in zones. In figure 8.1, the corresponding anatomical terms are indicated.

For space constraints we had excluded many maps from the paper. Here, we include additional maps for surface area, elongation, and water solubility, visible in fig. 8.2. It becomes clear that molecular length correlates highly with respect to co–activation with molecular elongation and surface area. Further, long molecules are less water soluble, in fact in the dataset, there is a negative correlation between

water solubility, on the one hand, and surface area, elongation, and water solubility, on the other hand. This could give rise to interpretations of areal specialization, along the lines of Wilson and Stevenson (2006, ch. 5), who observed that large, less volatile odorants provide relatively long–lasting and less easily detectable signals.

These results were based on the definition of representation and the statistical test thereof (c.f. section 7.3). It is important to note that we did not take into account correlations between the investigated physico–chemical properties.

We found several zones that are preferentially responsive for different combinations of properties. These properties included molecular bonds, cyclization, functional groups.

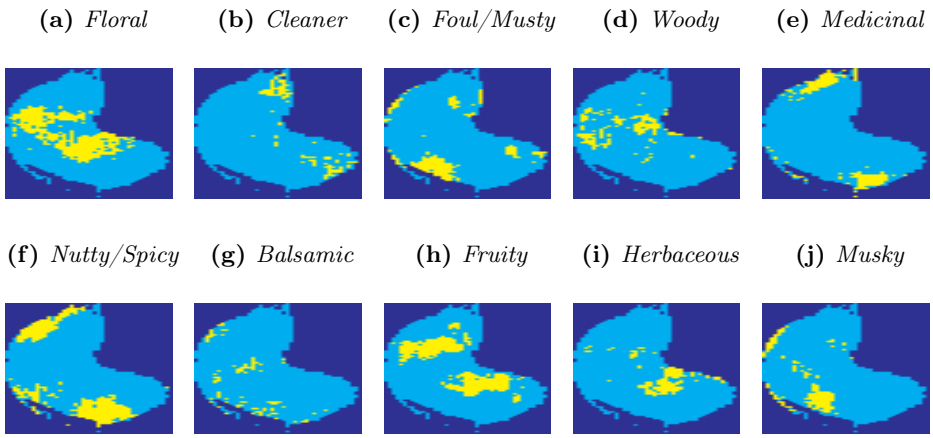
Further, by extension of our statistical procedure, we compared the size of coding zones. We are not aware of other attempts in the literature to systematically quantify the sizes of coding areas. It should be cautioned that these results should be interpreted more qualitatively than quantitatively.

We applied support–vector machine classification of physico–chemical odorant properties by glomerular activations in order to estimate their impact on early olfactory coding (this method is described in section 7.4). The logic behind is that properties that greatly change activations at the OB level should be easier to classify. Knowing the relevance of molecular properties could provide insights into the early coding of chemical information and provide vital clues for discerning which properties are functional in determining the degree of interaction between an OR and odorant molecules. We define relevance as the best classification performance from either the most representative points or a random baseline whichever was higher.

Our results indicate that some physico–chemical properties are especially relevant for glomerular representations. The properties with the highest performance scores were alkyne, alkane, alkene, and amine. This could indicate that these properties have a very strong impact on olfactory processing in rats. In general, these systematic results confirm earlier studies indicating the importance of cyclization, bond saturation, and some functional groups, for example Johnson and Leon (2007) and Yoshida and Mori (2007).

8.3 Results on perceptual representations in the rat olfactory bulb

Figure 8.4: ***Representations of perceptual categories.*** *Representations for odors (indicated in yellow) corresponding to the following perceptual categories:* ***a*** *— floral,* ***b*** *— cleaner,* ***c*** *— foul,* ***d*** *— woody,* ***e*** *— medicinal,* ***f*** *— nutty/spicy,* ***g*** *— balsamic,* ***h*** *— fruity,* ***i*** *— herbaceous, and* ***j*** *— musky.*



In paper V, we performed an analysis of coding for perceptual odor qualities. This was again done based on images from the glomerular response archive, which was described in section 7.1. The dataset of perceptual properties was described in section 7.2.

We started off by computing principal components and their correlations to odor qualities. These (linear) correlations indicated that some odor qualities aligned well along these principal components. As found in previous studies (e.g. Zarzo, 2008; Khan et al., 2007), the first dimension correlated with qualities that are known to be particularly pleasant and unpleasant, such as (pleasant) fruity and floral in one direction and (unpleasant) medicinal in the opposite direction. As stated, this reproduces similar results from the literature. However, we found the third principal dimension to show a similar alignment possibly related to pleasantness.

The main result from paper V was to find clustered responsive regions for some

odor qualities. These results are reproduced in figure 8.4. These odor qualities included floral, cleaner, foul, woody, medicinal, nutty, balsamic, fruity, herbaceous, and musk. To our knowledge, these results are novel.

We then compared the spatial proximity and overlap between these zones (measured by the Hausdorff distance) to perceptual distances as given in Zarzo (2008) as an evaluation if these distances could carry meaningful information. The theoretical aspect behind this was discussed before in chapter 6 in relation to chemical progression. As comparison to these spatial distances, for the population code, we compared dissimilarities based on the Euclidean distances between the (mean) population vectors and the perceptual distances (as per Zarzo 2008). The method for computing perceptual distances was explained in section 7.5.

Based on this comparison, we argued that spatial distances seemed to be more informative of perceptual differences than what could be expected from random distributions and better than distances between the population codes.

8.4 Results on organization of the olfactory bulb

In paper VI, we modeled the early olfactory system based on the methods described in section 7.6. The simulation included a dataset of odorants taken from Haddad et al. (2008a). This dataset contains ecologically–relevant smells in a high–dimensional space (32 dimensions) that had been optimized to fit with physiological data of rats, fruit flies, mice, honeybees, and tadpoles. In this space spanned by the odorants and their descriptors, we placed receptors using fuzzy clustering (the fuzzy c–means algorithm). These receptors were consequently located at partly overlapping regions in this space. The receptor response magnitude to an odorant depended on its distance to the odorant. We defined two different neural response functions. The first one is simplified to be ignorant to concentration or — in other words — describes the responses of ORNs with respect to a normalized concentration. The second response function depends on a concentration and was used to obtain dose–response curves.

We then simulated sensory responses to the set of odorants and defined regions of sensory axon projections based on the similarities between the neural response

profiles. The spatial organization was computed according to the method described in section 7.8, which included the computation of a distance matrix and dimensionality reduction by multi–dimensional scaling (c.f. section 7.7). The regions of synaptic convergence were computed using multi–dimensional scaling. At each iteration, odorants were presented to the network, distances between olfactory axons, and projections were adjusted. We demonstrated that in such a way, over iterations of the algorithm, axonal projections from ORNs of the same type converge at the bulb.

We next showed that according to our dose–response model, few glomeruli are activated at low concentration. As the concentration increases more glomeruli are activated, until the glomerular map becomes very unspecific and saturated. This means that our model accounts for increasing glomerular recruitment as reported in the literature (c.f. section 6.4; Rubin and Katz, 1999; Cleland et al., 2007; Johnson and Leon, 2000a; Stewart et al., 1979).

Then, similarly to paper V, we computed matches of spatial distances and population vectors, respectively, to perceptual differences (Zarzo, 2008) (dataset described in section 7.2). Table 8.1 reproduces these results. It shows that spatial distances explained perceptual distances much better than could population codes. Notably, we found significant (Spearman rank) correlations (at the 5% significance threshold) between the distances for one dataset and the spatial code (for the perfumers' dataset BP; $\rho = 0.6848, p = 0.04$), while other correlations were insignificant. The method for computing perceptual distance is explained in section 7.5.

Table 8.1: ***Matching error between perceptual spaces****. Absolute error of fit between coding spaces as computed from response patterns (D_P, D_S) and perceptual spaces (D_{Chrea} and $D_{BH\ small}$). Chrea and BH small are different perceptual spaces from the literature.*

	D_{Chrea}	$D_{\text{BH small}}$
D_S	**0.51**	**0.24**
D_P	0.60	0.55
combined	0.43	0.34
baseline	0.65	0.65

Chapter 9

Discussion

After we reviewed some background and literature in chapters 5 and 6, and introduced datasets and methods in chapter 7, we presented our results in chapter 8. This chapter discusses these results. In chapter 10, some conclusions will be drawn from them.

In this chapter, instead of dividing the work into sections by papers, we decided to follow a more problem–oriented structure in order to have a better reading flow.

For two of the studies, we used a dataset by the Michael Leon group (c.f. section 7.1). This dataset is publicly available on the internet and contains measurements of glomerular responses to odorants. We applied statistical methods and analysis based on machine learning techniques (compare section 7.3).

Some principle questions about early olfaction are the following:

- What is the organization of the sensory space?
- How is perceptual information encoded in the olfactory system and how can it be decoded?
  - Is there a spatial organization underlying activations in the olfactory system?
    - What are its principles?
    - What is its function?
  - What are the principles of projection to the glomerular layer?

The goal of our research into spatial coding in the glomerular layer of the olfactory bulb was to gain a better understanding about some of these questions. This includes issues about the organization of odor representations, coding dimensions in the olfactory system, the emergence of order based on self–organizing principles, information transfer in the brain, and the predictive value of activity maps in terms of perception.

This discussion starts off by the question of spatial organization and then comes to the other questions.

9.1 Is there a spatial organization underlying activations in the olfactory bulb?

As discussed earlier, representations in the brain have often been observed to be spatially organized at least in early stages and in the sensory cortices. For the perception of smells, the analysis of such an organization, however, is not as straightforward, because the physical metric is unclear (c.f. section 5.4.1). As discussed in section 6.4, there is accumulated evidence for glomerular feature clusters and some molecular properties are detected by topographically close glomeruli (e.g. Johnson et al., 2007; Mori et al., 2006b; Meister and Bonhoeffer, 2001; Rubin and Katz, 1999; Vassar et al., 1994). Many studies showed a receptive field organization on the level of glomeruli which correlates with evidence suggesting a pattern of input convergence of ORs onto topographically confined regions of the OB (for example Shepherd, 1972; Land and Shepherd, 1974; Lancet, 1986).

In chapter 7, we presented a nonparametric statistical procedure for decoding of continuous behavioral properties to investigate coding at the glomerular level of the OB. It was used in paper IV and paper V. It consisted of a univariate statistical test, the Wilcoxon rank–sum test, within a bootstrap wrapper. This procedure brings the advantage of systematic and quantitative results in the absence of assumptions of distributions and allows comparison of vectors of unequal lengths. By our method, we could map odorant properties to clustered zones. We found continuous spatial zones for a set of molecular properties and perceptual categories in papers IV and V, respectively. From these studies, we can draw conclusions about the spatial

organization of representations at the glomerular layer of the OB.

9.1.1 Spatial organization with respect to a subset of physico–chemical properties

It was shown earlier (Meister and Bonhoeffer, 2001; Rubin and Katz, 1999) that many glomeruli are sharply tuned for a small range of carbon chain lengths. Johnson and Leon (2007) found chemotopic progressions with increasing carbon number for several subgroups of odorants: (i) carboxylic acids, methyl and ethyl esters, (ii) primary alcohols, aldehydes, phenols, (iii) those with an aliphatic hydrocarbon chain. These progressions were shifts from medial and lateral areas into ventral areas.

Our results on the complete set of 362 odorants show mostly medial areas for up to 7 carbon atoms and ventral areas for 14–21, and ventro–medial areas for molecules of 8–13 carbon atoms. This confirms and extends Leon and Johnson's description of areas for molecular length. Leon and Johnson indicated that zones in dorsal areas were active for shorter chains, while longer molecules activated zones more ventrally. Our results for carboxylic acids coincide with Johnson and Leon (2007) in locating responsive glomeruli in anterior medial/lateral areas. We also confirm Leon and Johnson's result in placing regions for alcohols and phenols in lateral/medial areas, more caudal than regions for carboxylic acids. Further, we found areas for esters overlapping with their placement of aliphatic esters located in some central medial and lateral areas just next to alcohols. We found another area for esters in caudal areas of the lateral and medial bulb.

Johnson and Leon placed a zone for aromatics with oxygen groups and ketones in a dorsal region and one for aromatic hydrocarbons in a dorsal caudal region. We found a large region for aromatics, mostly dorsal but also medial and lateral, and a region for ketones in similar dorsal areas.

Soucy et al. (2009) had only found a coarse topography for similar odorants in mouse and rats, which could be taken as contrary to our results. We would caution, however, against a direct comparison, because of several methodological differences. (i) They used different imaging method (optical imaging vs. 2DG), which imply differences in temporal scale of measurements and therefore distortions of correlations

(compare Khan et al., 2010). (ii) There could also be differences in spatial scale between the datasets, since Soucy et al. did not average over multiple animals as did Leon and Johnson (c.f. section 7.1). Although, Soucy et al. report that there is little variation in location of glomeruli by response-profile across animals, this could influence results. (iii) Soucy et al. analyzed only a small part of the dorsal surface of the olfactory bulb, while Leon and Johnson's dataset provides response profiles over the whole spatial array of glomeruli. (iv) Soucy et al. investigated correlations of glomerular response profiles over a set of specific odorants, as opposed to correlations of responses to a set of properties of odorants. (v) Soucy and colleagues used anesthesized animals which may cause changes in dynamics which are only partially understood (compare Li et al., 2011; Khan et al., 2010).

In paper IV, we compared sizes of coding zones and relevance to representations between the investigated molecular properties. By application of a support vector machine classification procedure, as described in chapter 7, we obtained results for relevance of these properties. By extension of our statistical procedure, we compared the size of coding zones. We are not aware of other attempts in the literature to quantify sizes of coding areas. Results of size go in hand with localization maps shown above to indicate that some encoding of certain properties are specific to some spatial areas.

It should be cautioned, however, that these results should be interpreted more qualitatively than quantitatively. The thresholding of p–values at certain significance values means that effects of concentration and relevance cannot be completely separated, however, presented results could serve to roughly group properties by their coding zones. Size estimates of investigated properties differed largely. There are properties that recruit bigger zones and properties that recruit smaller zones. The properties for which we found the smallest coding zones are amine and sulfur–containing compound, with roughly 0.5% and 4.1% of recruited area.

It is known from many studies that the size of neural representation is related to perceptual acuity (e.g. Nienborg et al., 2004; Brown et al., 2004) and that exposure to olfactory stimuli leads to an increase of the responsive glomerular area (Rochefort et al., 2002; Woo and Leon, 1991). Therefore, we think that the area sizes we found could be related to acuity. Larger coding zones could mean that properties are

sensed by a broad range of ORs or that represented odorants have been presented frequently. In turn, it could be argued that properties with small coding zones and high relevance (as measured) could be more directly related to coding primitives, odotopes. From our results, amine, sulfur–containing compound, and alkyne could be such candidates.

Our classification results (see method in section 7.4) indicate that there are some properties that strongly affect odor coding on the OB level. Most relevant properties are alkyne, alkane, alkene, and amine. From our study, we speculate that these properties could have a big impact on olfactory processing in rats and we briefly discussed possible reasons for that.

Activations are very distinct with respect to whether an odorant contained a sulfur functional group or not. Bond saturation, indicative of the reactiveness of compounds, seems also to affect coding strongly, as we can see in the high classification performance of alkyne, alkane, and alkene. Carboxylic acid another functional group and aromatic, a cyclization property, still seemed to be quite important. Our results confirm that cyclization, bond saturation, and some functional groups are important. This is in line with Johnson and Leon (2007), who proposed as important dimensions of molecular properties cyclization, carbon numbers, bond saturation, branching, functional groups, and substitution position.

Our results also partly confirm Yoshida and Mori (2007), who proposed 14 primary odorant categories which could serve to enhance category–profile selectivity. These properties were sulfides, alcohols, methoxypyrazines, 6–carbon and 9–carbon green–odor compounds, aldehydes, ketones, isothiocyanates, terpene hydrocarbons, esters, terpene alcohols, alkylamines, acids, lactones, and phenol and its derivatives. As for the properties included in the study in paper IV, we found that, sulfides, alcohols–phenol, ketones, ester–lactone, amines, performance was good, however, our results indicate that other properties such as whether odorants contained a carboxylic–acid group or their bond saturation could also be important.

We find clustered glomerular representations for many molecular properties in a 2–deoxyglucose autoradiography data set of the rat OB. Of the compared molecular properties, presence of alkyne, alkane, alkene, and amine causes big changes in these activation maps as compared to other properties. Amines, sulfur–containing

compounds, and alkynes have small zones and high relevance to activation changes. Aromatics, alkanes, and carboxylic acids showed biggest zones. Together, these results suggest a local spatial encoding for some molecular properties.

It has to be mentioned that we did not control for correlations between molecular properties in our study. Long molecules are, for example, less water soluble. Areas for molecular length correlated highly with respect to co–activation with molecular elongation, surface area, and water solubility (not shown). It is known that carbon number is associated with molecular length, volume, hydrophobicity, among other properties (c.f. Johnson and Leon, 2007). In many studies only a small set of odorants is examined at differing concentration levels. A comparison across many studies is therefore not always straightforward, however, we think that our results are generally in good agreement with conclusions of several studies, including Johnson and Leon (2007).

9.2 How is perceptual information encoded in the olfactory bulb?

For the study in paper V we matched odor quality descriptors from Leon and Johnson (2006) to odor categories. Then we applied the statistical analysis to find representations. We found that representations of perceptual categories are typically arranged in continuous spatial zones. These results point towards a modular or domain organization. The implications for coding from such an organization are discussed in the paper and in section 9.2.1 in this thesis.

We identified representations for perceptual categories from rat glomerular response patterns to odors corresponding to these odor categories. This was done in two manners, according to population and according to spatial coding schemes. As for population coding, we took the strict interpretation of uniform weights for all units. As for spatial coding, we used a statistical method to calculate receptive fields corresponding to odor qualities for each glomerulus.

We first tried to relate principal components of glomerular response patterns to perceptual categories. Haddad et al. (2010) analyzed population activity of GLs, ORs, and secondary neurons from different studies and found that the first princi-

pal component was highly correlated with approach or withdrawal (or *pleasantness* vs. *unpleasantness*). Furthermore, they found that the first principal component correlated with many odor qualities.

In paper VI, we used the dataset from Haddad et al. (2008a) of realistic, biologically–relevant odorant data and showed how from these simple principles a topographic map emerges that is arranged by similarities defined by stimulus statistics. We then analyzed variability of glomerular responses over different odor categories. Data of odor categories were obtained by mapping odorant descriptors. We used a statistical method to calculate receptive fields corresponding to odor categories for each glomerulus. Then, we compared how well a spatial code and population code matched differences of human perceptual experiences. We found that errors of fit between spatial distances of representations, on the one hand, and perceptual orderings by human subjects on the other hand, were smaller than was the case for population codes. Therefore, our results suggest that spatial maps have a stronger relation to perception than have population codes. However, it has to be noted that also for population codes we obtain a better match than would be expected by chance. The combined codes integrating population and spatial codes, matched better in the case of the smaller set of odor descriptors (BH) and worse for Chrea. Therefore, overall from our data, it is not clear, whether population and spatial coding schemes complement each other.

In summary, our model shows that the spatial organization found for perceptual categories could be partly explained by regularities in the stimulus space.

9.2.1 What is the function of spatial organization?

As explained in section 9.2, our results suggest that there could be information about perceptual categories on spatial and population levels. Particularly, we found a local clustering of glomeruli, continuous spatial zones, for some molecular properties and for perceptual categories. This could point towards an organization into modules by perceptual relevance.

We already discussed some possible functional implications of this organization in section 6.4.1. As noted before, it can be speculated that the observed clustering of properties constitutes an instance of the minimization of wiring length in cortical

networks for local processing of feature combinations (c.f. Chen et al., 2006; Buzsáki et al., 2004). Spatial clustering could imply the existence of specialized filters and read–out mechanisms. As suggested before, clustering by similar sensitivity could allow a center–surround organization of lateral inhibitory filters.

Along this line, increasing recruitment of activated area could improve signal–to–noise ratio which allows a decrease of detection thresholds or it could indicate a further specialization within the recruited area. A smaller area could also indicate a specialized local feature detector. From that perspective, for example, our result that both amines and sulfur–containing compounds are represented in a small area of glomeruli could indicate the need for more thorough investigation.

It has been suggested before (e.g. Leon and Johnson, 2009) that spatial distances could be related to behavioral (or perceptual) differences. In order to test this hypothesis, we investigated relations of distances between regions that are preferentially active for perceptual categories to perceptual orderings. We compared how well a spatial code and population code matched human perceptual experiences. We found that for odor qualities many representations are spatially continuous, hence this speaks in favor of a spatial code.

We found that distances between spatial representations were similar to perceptual orderings by human subjects (as compared to baseline) and errors were smaller than for population codes. Therefore, our results suggest that spatial coding has a stronger relation to perception than population codes and confirm previous studies which suggest a spatial coding.

Cleland and Leon and Johnson argued based on experiments that qualitative odor perception is determined by glomerular activity patterns (Cleland et al., 2007; Johnson and Leon, 2000d). It is also known that activity patterns recruit more size by increasing odorant concentration (Khan et al., 2010; Johnson and Leon, 2000c; Sachse et al., 1999) and that perceptual qualities of odors can change depending on concentration (Wilson and Stevenson, 2006, ch. 4;e.g. Doty et al., 1975). In the spatial coding scheme, distances between patterns would change with increasing concentration and we suggest that such an effect could help explaining differences in discrimination ability.

9.3 What are the principles of projection to the glomerular layer?

As we discussed in section 6.4.1, it is not clear, if organizational principles behind the spatial arrangement of glomeruli can be pervasive in principle and if they could be useful for processing of smell information in the brain, given that smell is such a high–dimensional stimulus (c.f. Wilson and Stevenson, 2006; Cleland and Sethupathy, 2006). We think that a better understanding of the olfactory stimulus is needed in order to make sense of the organization of representations.

In paper VI, we modeled the axon growth of sensory neurons based on co–variance of neurons. As stated earlier and explained in section 7.6, this was done based on a dataset of 447 odorants, each of which described by 32 physico–chemical descriptors. The choice of descriptors maximized correlations between variability of OB/antennal lobe activity responses, on the one hand, and variability of the odorant descriptors, on the other hand. In our abstract model, we placed OR receptive fields in the space captured by these molecular odorant descriptors and defined positions of ORN axon projections onto the bulb by dimensionality reduction of the OR–odorant affinity correlation matrix. Our model included a population mean rate in response to a concentration. In order to establish the plausibility of this axon growth mechanism, apart from referring to biological literature, we showed how activity–dependent mechanisms could serve to organize olfactory axons into glomerular structures by OR identity. Axons of receptor neurons of different types are intermixed as they grow toward the brain and they have to undergo a sorting process before arriving at their target glomeruli. In our model, olfactory axons clustered together over several iterations of the MDS algorithm. We then showed how glomerular responses spatially broadened and saturated with increasing concentrations. It is well–established that increasing concentration leads to an increasing recruitment of glomeruli and thereby to a spatial broadening of local peaks (Khan et al., 2010; Johnson and Leon, 2000c; Sachse et al., 1999).

We showed in paper paper VI that a model of cortical projections (Lansner et al., 2009) can be extended to explain the emergence of topography from statistics of naturally occurring odors (as discussed in section 7.6). The principle of axonal

coalescence by co–activation gives rise to a topographic map where the distance of the components in the representation is related to perceptual difference. We showed how the activity–dependent axon growth could lead to the formation of glomeruli, the emergence of spatial concentration coding, and topographically organized receptive fields. We showed that there is a relationship between spatial separation of coding zones in the model and perceptual differences in humans. This could indicate that in theory such a pervasive topographic organization is indeed possible.

9.4 What is the organization of the sensory space?

We hope that our model can provide insights into the formation of the olfactory bulb map and internalizations of environmental regularities, even though the match between glomerular activity and human perceptual space could be improved by tuning parameters of the receptor affinity distribution or by including top–down projections. Other factors influencing axon guidance are also known and the balance between the different factors is unclear.

As for the placement of receptors (profiles of virtual receptors), it could be that odorant receptors and projections to the OB are optimized to make environmental regularities more prominent. For example, Geisler and Diehl (2002) proposed that the design of perceptual systems is optimized according to statistics of natural stimuli and evolutionary fitness. More particular for olfaction, Nei et al. (2008) discussed that part of molecular changes in chemoreceptor repertoires constitutes an adaptation of organisms to different environments. On the other side, it was suggested (Abbott and Luo, 2007) that olfactory receptors are not optimal either and Sánchez-Montañés and Pearce (2002) demonstrated that optimal stimulus estimation arises from a local randomized mechanism for receptor specificity generation. Thus, while our virtual receptors lack a direct biological correspondence, the principles of their placement seem plausible and reasonable within the scope of our model.

Chapter 10

Conclusions

In chapters 8 and 9, we summarized and discussed results regarding the processing of odors in the early olfactory pathway of vertebrates. In this chapter, we will draw conclusions from the presented studies.

In paper IV, we investigated specifically these questions: which odorant properties are coded and where, what is the size of the coding zones, and how relevant are individual odorant properties to the encoding.

We performed a systematic study on several hundred odorants and more than a dozen of physico–chemical properties to investigate the coding for physico–chemical properties at the OB level, based on a dataset of responses to more than 360 odorants. Statistical analysis and machine learning techniques were used for this purpose. By statistical analysis, we identified sites of representations for many odorant properties. We systematically determined that representations for many properties, such as functional groups and properties of molecular shape. Generally, we found that many properties are coded in specific areas in the OB and that the extent of these zones tends to be rather limited. In a classification study, we found that the spatial patterns of glomerular activation are indicative of odor classes and that the prediction of physico–chemical properties based on small areas was already very reliable. We confirmed a spatial progression of representations by carbon–chain length. We compared the relevance of some physico–chemical properties for these representations using classification performance and found differences with respect to how strongly properties affect representations at the OB level. We discussed these

differences in the context of sensitivity enhancement and read–out mechanisms.

Based on the same dataset, we extracted perceptual categories of odors from the literature and computed spatial codes from rat glomerular response patterns for these categories. We first analyzed the relationships between the principal components of the glomerular response patterns and perceptual categories and roughly confirmed previous findings about population codes at the OB. We found that representations of perceptual categories are confined to continuous spatial zones. Further, we compared distances between these zones to human perceptual differences. We found that in general the main orientations of perceptual and coding spaces matched, which suggests that perceptual qualities could be represented, to some degree, in codes at the level of the first sensory projection. We found that the distances between local coding zones span a space that is similar to human–subject perceptual ordering, and that the distances between spatial representations are more accurate in predicting perceptual differences than distances between population codes. This suggests that these spatial continuous representations are organized in a meaningful way in terms of perceptual relationships.

We built a model of the early olfactory system where we generated virtual receptors and based sorting of sensory neuron projections on the correlated neural activities. We used realistic, biologically–relevant odorant data and showed that this model reproduces basic phenomenological features, such as glomerular clustering by OR identity and glomerular changes with respect to concentration changes. We found that spatial distances based predicted by topography explained perceptual differences better than distances predicted by population coding. The model showed how natural statistics of odors can give rise to an ordered arrangement of representations at the OB level where distances between representations are perceptually relevant. Therefore, these findings support our previous study which suggested a spatial coding for perceptual categories at the OB. Generally, our results from paper VI point to relationships between natural olfactory stimulus characteristics and perceptual dimensions.

Although the machinery responsible for axon guidance is much more complex than that presented in paper VI, we hope that progress of functional understanding may be facilitated by keeping our model as simple as possible. The formation of

topographic organization in our model relies exclusively on the input side, however, other factors could be integrated in our model.

Together, the presented results suggest the existence of spatial receptive fields at the glomerular level and that the relative spatial arrangement of these receptive fields could be meaningful in terms of functionality and help to organize stimuli based on perceptual categories. The question remains about functional implications of this spatial organization. As we discussed, it has been suggested that the topographic glomerular organization could facilitate a decrease of detection limits (by increase of signal–to–noise ratio) and a sharpening of inputs over periglomerular connections, and we hypothesize that this organization allows for higher–order feature detection at the level of seconday neurons. Our results could indicate organizing principles that could serve to efficiently convey behaviorally relevant information to higher stages, e.g. to the amygdala over slow and fast routes. Thus, we conjecture that a read–out mechanism could sample from glomeruli in spatial domains to extract behaviorally relevant information from stimuli.

As for future work, a lot needs to be done. The results from these studies still need to be confirmed using different recording techniques. The characterization of physiological responses needs to be extended in terms of physico–chemical and perceptual properties. Normative results such as from paper VI can be extended to study theoretical aspects of coding. Different model studies could investigate functionality of OB and downstream circuitry.

List of Figures

References

T. Abaffy and A. R. Defazio. The location of olfactory receptors within olfactory epithelium is independent of odorant volatility and solubility. *BMC research notes*, 4(1):137, January 2011. ISSN 1756-0500. [102]

L. F. Abbott and S. X. Luo. A step toward optimal coding in olfaction. *Nature neuroscience*, 10(11):1342–3, November 2007. ISSN 1097-6256. [164]

H. Abdi. *Les réseaux de neurones*. Presses universitaires de Grenoble, 1994. [47]

H. Abe, S. Kanaya, T. Komukai, Y. Takahashi, and S.-i. Sasaki. Systemization of semantic descriptions of odors. *Analytica Chimica Acta*, 239:73–85, 1990. ISSN 00032670. [116]

N. M. Abraham, V. Egger, D. R. Shimshek, R. Renden, I. Fukunaga, R. Sprengel, P. H. Seeburg, M. Klugmann, T. W. Margrie, A. T. Schaefer, and T. Kuner. Synaptic inhibition in the olfactory bulb accelerates odor discrimination in mice. *Neuron*, 65(3):399–411, February 2010. ISSN 1097-4199. [107]

B. W. Ache and J. M. Young. Olfaction: diverse species, conserved principles. *Neuron*, 48 (3):417–30, November 2005. ISSN 0896-6273. [93]

A. Acik. *Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories.* masters, Osnabrueck, June 2006. URL `http://www.ncbi.nlm.nih.gov/pubmed/19306892`. [54]

T. E. Acree and H. Arn. Flavornet: A database of aroma compounds based on odor potency in natural products. In E. Contis, C.-T. Ho, C. Mussinan, T. Parliment, F. Shahidi, and A. Spanie, editors, *Food Flavors: Formation, Analysis and Packaging Influences, Proceedings of the 9th International Flavor Conference The George Charalambous Memorial Symposium*, volume 40, page 27. Elsevier, 1998. ISBN 0167-4501. [115, 136]

Y. Adam and A. Mizrahi. Circuit formation and maintenance–perspectives from the mammalian olfactory bulb. *Current opinion in neurobiology*, 20(1):134–40, February 2010. ISSN 1873-6882. [132]

J. Albrecht, M. Demmel, V. Schöpf, A. M. Kleemann, R. Kopietz, J. May, T. Schreder, R. Zernecke, H. Brückmann, and M. Wiesmann. Smelling chemosensory signals of males

in anxious versus nonanxious condition increases state anxiety of female subjects. *Chemical senses*, 36(1):19–27, January 2011. ISSN 1464-3553. [95]

R. Aloni, T. Olender, and D. Lancet. Ancient genomic architecture for mammalian olfactory receptor clusters. *Genome biology*, 7(10):R88, January 2006. ISSN 1465-6914. [92, 99, 102]

J.-M. Alonso and Y. Chen. Receptive field. *Scholarpedia*, 4(1):5393, 2009. ISSN 1941-6016. [125]

H. Alt, C. Knauer, and C. Wenk. Comparison of Distance Measures for Planar Curves. *Algorithmica*, 38(1):45–58, October 2003. ISSN 0178-4617. [142]

L. Altenberg. The Schema Theorem and Price âĂŹ s Theorem. In *Foundations of genetic algorithms*, volume 3, pages 23–49. Citeseer, 1995. [87]

O. Alter, P. O. Brown, and D. Botstein. Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 97(18):10101–10106, 2000. [84]

J. E. Amoore. Specific Anosmia and the Concept of Primary Odors. *Chemical Senses*, 2 (3):267–281, 1977. ISSN 0379-864X. [116]

A. K. Anderson, K. Christoff, I. Stappen, D. Panitz, D. G. Ghahremani, G. Glover, J. D. E. Gabrieli, and N. Sobel. Dissociated neural representations of intensity and valence in human olfaction. *Nature neuroscience*, 6(2):196–202, February 2003. ISSN 1097-6256. [110]

J. Antolík and J. A. Bednar. Development of maps of simple and complex cells in the primary visual cortex. *Frontiers in computational neuroscience*, 5(April):17, January 2011. ISSN 1662-5188. [106]

A. Arzi and N. Sobel. Olfactory perception as a compass for olfactory neural maps. *Trends in cognitive sciences*, pages 1–9, October 2011. ISSN 1879-307X. [115, 125]

K. Asahina, V. Pavlenkovich, and L. B. Vosshall. The survival advantage of olfaction in a competitive environment. *Current biology : CB*, 18(15):1153–5, August 2008. ISSN 0960-9822. [92]

L. Astic and D. Saucier. Anatomical mapping of the neuroepithelial projection to the olfactory bulb in the rat. *Brain Research Bulletin*, 16(4):445–454, April 1986. ISSN 03619230. [131]

B. Auffarth. Spectral Graph Clustering. Technical report, Universitat de Barcelona, 2007. URL `http://www.csc.kth.se/~auffarth/publications/Auffarth_SpectralGraphClustering_2007.pdf`. [15]

B. Auffarth. How to Build a Linux Cluster for Scientific Computing. Technical report, Universitat de Barcelona, 2009. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.4705&rep=rep1&type=pdf.http://www.csc.kth.se/~auffarth/publications/beowulf.pdf`. [140]

B. Auffarth. Clustering by a genetic algorithm with biased mutation operator. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE, July 2010. ISBN 978-1-4244-6909-3. [5]

B. Auffarth, M. López, and J. Cerquides. Hopfield networks in relevance and redundancy feature selection applied to classification of biomedical high-resolution micro-CT images. In *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, pages 16–31. Springer, 2008. [4]

B. Auffarth, M. Lopez, and J. Cerquides. Comparison of Redundancy and Relevance Measures for Feature Selection in Tissue Classification of CT images. In P. Perner, editor, *Advances in Data Mining - Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, pages 248–262. Springer Heidelberg, lnai 6171 edition, 2010. [4]

B. Auffarth, G.-G. Galvez, and S. Marco. Continuous spatial representations in the olfactory bulb may reflect perceptual categories. *Frontiers in Systems Neuroscience*, 5(0):1–14, 2011a. [5, 126, 132]

B. Auffarth, A. Gutierrez-Galvez, and S. Marco. Statistical analysis of coding for molecular properties in the olfactory bulb. *Frontiers in systems neuroscience*, 5(0):62, January 2011b. ISSN 1662-5137. [5, 126, 128, 129, 130, 131]

B. Auffarth, B. Kaplan, and A. Lansner. Map formation in the olfactory bulb by axon guidance of olfactory neurons. *Frontiers in Systems Neuroscience*, 5(0):1–16, 2011c. [5, 98, 119, 132]

J. L. Aungst, P. M. Heyward, a. C. Puche, S. V. Karnup, A. Hayar, G. Szabo, and M. T. Shipley. Centre-surround inhibition among olfactory bulb glomeruli. *Nature*, 426(6967): 623–9, December 2003. ISSN 1476-4687. [129]

B. B. Averbeck and M. Seo. The statistical neuroanatomy of frontal networks in the macaque. *PLoS computational biology*, 4(4):e1000050, April 2008. ISSN 1553-7358. [101]

R. Axel. The Molecular Logic Of Smell. *Scientific American*, 16(3):68–75, September 2006. ISSN 0036-8733. [115]

S. Ayabe-Kanamura, I. Schicker, M. Laska, R. Hudson, H. Distel, T. Kobayakawa, and S. Saito. Differences in perception of everyday odors: a Japanese-German cross-cultural study. *Chemical senses*, 23(1):31–8, February 1998. ISSN 0379-864X. [116]

H. B. Barlow. The Ferrier Lecture, 1980: Critical Limiting Factors in the Design of the Eye and Visual Cortex. *Proceedings of the Royal Society B: Biological Sciences*, 212(1186): 1–34, May 1981. ISSN 0962-8452. [129]

R. A. Baron. The Sweet Smell of... Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls. *Personality and Social Psychology Bulletin*, 23(5): 498–503, May 1997. ISSN 0146-1672. [97]

R. A. Barton, a. Purvis, and P. H. Harvey. Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 348(1326):381–92, June 1995. ISSN 0962-8436. [91, 92]

B. Bathellier, O. Gschwend, and A. Carleton. Temporal Coding in Olfaction. In A. Menini, editor, *The Neurobiology of Olfaction*, chapter 13, pages 329–351. CRC Press, Boca Raton (FL), 2010. [120]

G. Beni. A least biased fuzzy clustering method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):954–960, 1994. ISSN 01628828. [16]

A. Z. Berna, A. R. Anderson, and S. C. Trowell. Bio-benchmarking of electronic nose sensors. *PloS one*, 4(7):e6406, January 2009. ISSN 1932-6203. [96]

K. Berridge and J. Fentress. Trigeminal-taste interaction in palatability processing. *Science*, 228(4700):747–750, May 1985. ISSN 0036-8075. [95]

J. Beshel, N. Kopell, and L. M. Kay. Olfactory bulb gamma oscillations are enhanced with task demands. *Journal of Neuroscience*, 27(31):8358–65, 2007. ISSN 15292401. [120]

J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function*. Plenum Press, 1981. ISBN 0306406713. [14, 144]

J. C. Bezdek, R. Ehrlich, and W. Full. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984. ISSN 00983004. [85]

K. P. Bhatnagar, R. C. Kennedy, G. Baron, and R. A. Greenberg. Number of mitral cells and the bulb volume in the aging human olfactory bulb: a quantitative morphological study. *The Anatomical record*, 218(1):73–87, May 1987. ISSN 0003-276X. [99]

I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–117, 1987. ISSN 0033-295X. [50]

J. Biesiada and W. Duch. Feature selection for high-dimensional data: A kolmogorov-smirnov correlation-based filter. In *Computer Recognition Systems*, number 30/2005, pages 95–103. Springer, 2005. [25, 29, 38, 44, 45]

H. Boelens and H. Haring. Molecular structure and olfactive quality. Technical report, Naarden International, Bussum, the Netherlands, 1981. [141]

T. Bozza, P. Feinstein, C. Zheng, and P. Mombaerts. Odorant receptor expression defines functional units in the mouse olfactory system. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(8):3033–43, April 2002. ISSN 1529-2401. [98, 102]

A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, July 1997. ISSN 00313203. [141]

R. Briandet, E. K. Kemsley, and R. H. Wilson. Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. *Journal of Agricultural and Food Chemistry*, 44(1):170–174, January 1996. ISSN 0021-8561. [112]

P. B. Brown, H. R. Koerber, and R. Millecchia. From innervation density to tactile acuity: 1. Spatial representation. *Brain research*, 1011(1):14–32, June 2004. ISSN 0006-8993. [158]

L. B. Buck and R. Axel. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1):175–87, April 1991. ISSN 0092-8674. [100]

K. Bunte, M. Biehl, and B. Hammer. Dimensionality reduction mappings. In *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 349–356. IEEE, April 2011. ISBN 978-1-4244-9926-7. [145]

P. Burt and E. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983. ISSN 0096-2244. [52, 53]

D. Buschhüter, M. Smitka, S. Puschmann, J. C. Gerber, M. Witt, N. D. Abolmaali, and T. Hummel. Correlation between olfactory bulb volume and olfactory function. *NeuroImage*, 42(2):498–502, August 2008. ISSN 1095-9572. [92, 102]

a. J. Butte and I. S. Kohane. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 418–29, January 2000. ISSN 1793-5091. [16]

M. Butz, F. Wörgötter, and A. van Ooyen. Activity-dependent structural plasticity. *Brain research reviews*, 60(2):287–305, May 2009. ISSN 0165-0173. [105]

G. Buzsáki, C. Geisler, D. a. Henze, and X.-J. Wang. Interneuron Diversity series: Circuit complexity and axon wiring economy of cortical interneurons. *Trends in neurosciences*, 27(4):186–93, April 2004. ISSN 0166-2236. [129, 162]

W. S. Cain and F. Johnson. Lability of odor pleasantness: influence of mere exposure. *Perception*, 7(4):459–465, 1978. ISSN 0301-0066. [114]

T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics Theory and Methods*, 3(1):1–27, 1974. ISSN 03610926. [17]

P. Callegari, J. Rouault, and P. Laffort. Olfactory quality: from descriptor profiles to similarities. *Chemical senses*, 22(1):1–8, February 1997. ISSN 0379-864X. [115]

L. Cardamone, A. Laio, V. Torre, R. Shahapure, and A. Desimone. Cytoskeletal actin networks in motile cells are critically self-organized systems synchronized by mechanical interactions. *Proceedings of the National Academy of Sciences of the United States of America*, August 2011. ISSN 1091-6490. [105]

R. Caruana and A. Niculescu-Mizil. Data mining in metric space. In *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, page 69, New York, New York, USA, 2004. ACM Press. ISBN 1581138889. [22]

L. J. Cauller. Layer I of primary sensory neocortex: where top-down converges upon bottom-up. *Behavioural brain research*, 71(1-2):163–70, November 1995. ISSN 0166-4328. [120]

J. Cerquides, M. Lopez-Sanchez, S. Ontanon, E. Puertas, A. Puig, O. Pujol, and D. Tost. Learning Methods for Automatic Classification of Biomedical Volume Datasets. In *Conferencia de la Asociacion Espanola para la Inteligencia Artificial*, volume D, 2005. [23, 52]

H. D. Chalke, J. R. Dewhurst, and C. W. Ward. Loss of sense of smell in old people: a possible contributory factor in accidental poisoning from town gas. *Public health*, 72(6): 223–30, September 1958. ISSN 0033-3506. [96]

H.-P. Chan, L. Hadjiiski, C. Zhou, and B. Sahiner. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography-a review. *Academic radiology*, 15(5): 535–55, May 2008. ISSN 1076-6332. [52]

J. Chandrashekar, M. a. Hoon, N. J. P. Ryba, and C. S. Zuker. The receptors and cells for mammalian taste. *Nature*, 444(7117):288–94, November 2006. ISSN 1476-4687. [128]

C.-c. Chang and C.-j. Lin. LIBSVM: a library for support vector machines. *Computer*, pages 1–30, 2001. [56, 140]

M. Chastrette. Data management in olfaction studies. *SAR and QSAR in Environmental Research*, 8(3-4):157–181, 1998. [115]

M. Chastrette, A. Elmouaffek, and P. Sauvegrain. A multidimensional statistical study of similarities between 74 notes used in perfumery. *Chemical Senses*, 13(2):295–305, 1988. ISSN 0379-864X. [116]

S. Chen and P. S. Buckmaster. Stereological analysis of forebrain regions in kainate-treated epileptic rats. *Brain research*, 1057(1-2):141–52, September 2005. ISSN 0006-8993. [99]

W. R. Chen and G. M. Shepherd. The olfactory glomerulus: a cortical module with specific functions. *Journal of neurocytology*, 34(3-5):353–60, September 2005. ISSN 0300-4864. [98, 102, 104]

X. Chen, M. Gabitto, Y. Peng, N. J. P. Ryba, and C. S. Zuker. A Gustotopic Map of Taste Qualities in the Mammalian Brain. *Science*, 333(6047):1262–1266, September 2011a. ISSN 0036-8075. [127, 128]

X. Chen, X. Zhou, and S. T. C. Wong. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Transactions on Biomedical Engineering*, 53(4):762–766, 2006. [129, 162]

Y.-c. Chen, D. Mishra, L. Schmitt, M. Schmuker, and B. Gerber. A behavioral odor similarity "space" in larval Drosophila. *Chemical senses*, 36(3):237–49, March 2011b. ISSN 1464-3553. [114]

M. R. Chernick. *Bootstrap Methods: A Guide for Practitioners and Researchers*, volume 619 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, 2 edition, 2011. ISBN 1118211596. [139]

J. H. Cho, J. E. a. Prince, and J.-F. Cloutier. Axon guidance events in the wiring of the mammalian olfactory system. *Molecular neurobiology*, 39(1):1–9, February 2009. ISSN 0893-7648. [105, 106]

G. Choi, D. Stettler, B. Kallman, S. Bhaskar, A. Fleischmann, and R. Axel. Driving Opposing Behaviors with Ensembles of Piriform Neurons. *Cell*, 146(6):1004–1015, September 2011. ISSN 00928674. [111]

C. Chrea. Culture and odor categorization: agreement between cultures depends upon the odors. *Food Quality and Preference*, 15(7-8):669–679, December 2004. ISSN 09503293. [116, 141]

C. Chrea, D. Valentin, C. Sulmont-Rossé, D. H. Nguyen, and H. Abdi. Semantic, typicality and odor representation: a cross-cultural study. *Chemical senses*, 30(1):37–49, January 2005. ISSN 0379-864X. [116]

J. M. Christie, C. Bark, S. G. Hormuzdi, I. Helbig, H. Monyer, and G. L. Westbrook. Connexin36 mediates spike synchrony in olfactory bulb glomeruli. *Neuron*, 46(5):761–72, June 2005. ISSN 0896-6273. [99]

A. M. Cleary, K. E. Konkel, J. S. Nomi, and D. P. McCabe. Odor recognition without identification. *Memory & cognition*, 38(4):452–60, June 2010. ISSN 1532-5946. [93]

T. A. Cleland. The construction of olfactory representations. In *Mechanisms of information processing in the brain: encoding of information in neural populations.*, pages 1–59. Cambridge University Press, Cambridge, UK, 2008. ISBN 1775254275. [97]

T. A. Cleland and P. Sethupathy. Non-topographical contrast enhancement in the olfactory bulb. *BMC Neuroscience*, 7(1):7, 2006. [128, 163]

T. a. Cleland, A. Morse, E. L. Yue, and C. Linster. Behavioral models of odor similarity. *Behavioral neuroscience*, 116(2):222–31, April 2002. ISSN 0735-7044. [130, 131, 132]

T. A. Cleland, B. A. Johnson, M. I. Leon, and C. Linster. Relational representation in the olfactory system. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1953–8, February 2007. ISSN 0027-8424. [102, 123, 124, 142, 154, 162]

T. a. Cleland, S.-Y. T. Chen, K. W. Hozer, H. N. Ukatu, K. J. Wong, and F. Zheng. Sequential mechanisms underlying concentration invariance in biological olfaction. *Frontiers in Neuroengineering*, 4(January):1–12, 2012. ISSN 1662-6443. [107]

P. J. Clyne, C. G. Warr, M. R. Freeman, D. Lessing, J. Kim, and J. R. Carlson. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila. *Neuron*, 22(2):327–38, February 1999. ISSN 0896-6273. [99]

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995. ISSN 0885-6125. [56, 140]

T. Cover. The best two independent measurements are not the two best. *Systems, Man and Cybernetics, IEEE Transactions on*, 1(1):116–117, 1974. [25, 28]

C. J. Crasto. Computational Biology of Olfactory Receptors. *Current bioinformatics*, 4(1): 8–15, January 2009. ISSN 1574-8936. [112]

E. Crocker and L. Henderson. Analysis and classification of odors. *American Perfumer and Essential Oil Review*, 22:325—-327, 1927. [116]

J. L. Crowley and R. M. Stern. Fast computation of the difference of low-pass transform. *IEEE transactions on pattern analysis and machine intelligence*, 6(2):212–22, February 1984. ISSN 0162-8828. [52]

D. J. Cunningham. *Cunningham's textbook of anatomy.* William Wood and company, New York, New York, USA, 1918. [101, 103]

R. I. Damper and S. R. Harnad. Neural network models of categorical perception. *Perception & psychophysics*, 62(4):843–67, May 2000. ISSN 0031-5117. [122]

O. Danielsson and S. Carlsson. Generic Object Class Detection using Feature Maps. In *Image Analysis*, volume 6688/2011 of *Lecture Notes in Computer Science*, pages 348–359. 2011. [52]

D. L. Davies and D. W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, April 1979. ISSN 0162-8828. [17]

I. G. Davison and M. D. Ehlers. Neural circuit mechanisms for pattern detection and feature combination in olfactory cortex. *Neuron*, 70(1):82–94, April 2011. ISSN 1097-4199. [111]

D. R. Dawson and H. P. Killackey. The organization and mutability of the forepaw and hindpaw representations in the somatosensory cortex of the neonatal rat. *The Journal of comparative neurology*, 256(2):246–56, February 1987. ISSN 0021-9967. [128]

F. de Castro. Wiring Olfaction: The Cellular and Molecular Mechanisms that Guide the Development of Synaptic Connections from the Nose to the Cortex. *Frontiers in neuroscience*, 3(December):52, January 2009. ISSN 1662-453X. [101]

J. de Olmos, H. Hardy, and L. Heimer. The afferent connections of the main and the accessory olfactory bulb formations in the rat: an experimental HRP-study. *The Journal of comparative neurology*, 181(2):213–44, September 1978. ISSN 0021-9967. [109]

M. Dehnhard. Mammal semiochemicals: understanding pheromones and signature mixtures for better zoo-animal husbandry and conservation. *International Zoo Yearbook*, 45(1): 55–79, January 2011. ISSN 00749664. [93]

J. Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7(December):1–30, 2006. [58, 61]

L. Denoeud, H. Garreta, and A. Gu. Comparison of distance indices between partitions. In P. L., editor, *Applied Stochastic Models and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 21–28. Springer Berlin Heidelberg, 2005. ISBN 9783540344162. [17]

R. a. Dielenberg and I. S. McGregor. Defensive behavior in rats towards predatory odors: a review. *Neuroscience and biobehavioral reviews*, 25(7-8):597–609, December 2001. ISSN 0149-7634. [132]

C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2):185–205, 2005. [25, 43]

J. Djordjevic, M. Jones-Gotman, K. De Sousa, and H. Chertkow. Olfaction in patients with mild cognitive impairment and Alzheimer's disease. *Neurobiology of aging*, 29(5): 693–706, May 2008. ISSN 1558-1497. [96]

B. Doerr and A. Auger. *Theory of Randomized Search Heuristics: Foundations and Recent Developments*. Series on Theoretical Computer Science. World Scientific Publishing Co Pte Ltd, 2011. ISBN 9814282669. [18]

B. J. Doleman. Trends in odor intensity for human and electronic noses: Relative roles of odorant vapor pressure vs. molecularly specific odorant binding. *Proceedings of the National Academy of Sciences*, 95(10):5442–5447, May 1998. ISSN 00278424. [114]

L. Donna. Fragrance Perception: Is Everything Relative? *Perfumer & flavorist*, 34(12), 2009. [116]

K. M. Dorries, H. J. Schmidt, G. K. Beauchamp, and C. J. Wysocki. Changes in sensitivity to the odor of androstenone during adolescence. *Developmental psychobiology*, 22(5): 423–35, July 1989. ISSN 0012-1630. [94]

R. L. Doty. The olfactory system and its disorders. *Seminars in neurology*, 29(1):74–81, March 2009. ISSN 0271-8235. [96]

R. Doty, M. Ford, G. Preti, and G. Huggins. Changes in the intensity and pleasantness of human vaginal odors during the menstrual cycle. *Science*, 190(4221):1316–1318, December 1975. ISSN 0036-8075. [162]

J. Dougherty, R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. *Machine Learning: Proceedings of the Twelfth International Conference*, 54(2):194–202, 1995. [77]

A. Dravnieks. *Atlas of odor character profiles*. Astm Intl, 1985. ISBN 0803104561. [115]

L. Dryer. Evolution of odorant receptors. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 22(9):803–10, September 2000. ISSN 0265-9247. [91]

C. Dubacq, S. Jamet, and A. Trembleau. Evidence for developmentally regulated local translation of odorant receptor mRNAs in the axons of olfactory sensory neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(33): 10184–90, August 2009. ISSN 1529-2401. [106]

M.-P. Dubuisson and A. Jain. A modified Hausdorff distance for object matching. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, pages 566–568, Jerusalem, Israel, 1994. IEEE. ISBN 0-8186-6265-4. [142]

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2 edition, 2000. ISBN 0471056693. [11, 33, 34]

C. Dulac and a. T. Torello. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature reviews. Neuroscience*, 4(7):551–62, July 2003. ISSN 1471-003X. [94]

W. Dwinnell. SampleError function for Matlab. URL `http://matlabdatamining.blogspot.com/2007/06/roc-curves-and-auc.html`. [58]

B. Efron. The jackknife, the bootstrap and other resampling plans. In *CMBS Regional Conference Series in Applied Mathematics*, page 92. Philadelphia: Society for Industrial and Applied Mathematics, 1982. ISBN 0898711797. [139]

W. Einhäuser, W. Kruse, K.-P. Hoffmann, and P. König. Differences of monkey and human overt attention under natural conditions. *Vision research*, 46(8-9):1194–209, April 2006. ISSN 0042-6989. [54]

H. L. Eisthen. Evolution of Vertebrate Olfactory Systems. *Brain, Behavior and Evolution*, 50(4):222–233, 1997. ISSN 1421-9743. [91]

H. L. Eisthen. Why are olfactory systems of different animals so similar? *Brain, behavior and evolution*, 59(5-6):273–93, January 2002. ISSN 0006-8977. [93]

R. Elsaesser and J. Paysan. The sense of smell, its signalling pathways, and the dichotomy of cilia and microvilli in olfactory sensory cells. *BMC neuroscience*, 8 Suppl 3:S1, January 2007. ISSN 1471-2202. [103]

J. Esslen and K.-e. Kaissling. Zahl und Verteilung antennaler Sensillen bei der Honigbiene (Apis mellifera L .). *Zoomorphologie*, 251:227–251, 1976. [99]

A. Fantana. *Odor processing in the olfactory bulb: Structure of mitral cell receptive fields*. PhD thesis, Harvard University, 2006. URL `http://gradworks.umi.com/32/17/3217724.htmlhttp://focus.lib.kth.se/docview/305338225?accountid=7414`. [102, 134]

A. L. Fantana, E. R. Soucy, and M. Meister. Rat olfactory bulb mitral cells receive sparse glomerular inputs. *Neuron*, 59(5):802–814, September 2008. ISSN 1097-4199. [108]

T. Fawcett. ROC Graphs : Notes and Practical Considerations for Researchers. *ReCALL*, 31(HPL-2003-4):1–38, 2004. ISSN 08997667. [22, 23]

C. Ferri, J. Hernandez-Orallo, and R. Modroiu. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1):27–38, January 2009. ISSN 01678655. [22]

A. Fiala, T. Spall, S. Diegelmann, B. Eisermann, S. Sachse, J.-M. Devaud, E. Buchner, and C. G. Galizia. Genetically expressed cameleon in Drosophila melanogaster is used to visualize olfactory information in projection neurons. *Current biology : CB*, 12(21): 1877–84, October 2002. ISSN 0960-9822. [109]

M. Filippone, F. Camastra, F. Masulli, and S. Rovetta. A survey of kernel and spectral methods for clustering. *Pattern Recognition*, 41(1):176–190, January 2008. ISSN 00313203. [15]

H. T. Finck. The gastronomic value of odours. In *Food and flavor: a gastronomic guide to health and good living*, pages 559–581. The Century Co., 1913. [95]

M. L. Fletcher, A. M. Smith, A. R. Best, and D. A. Wilson. High-frequency oscillations are not necessary for simple olfactory discriminations in young rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(4):792–8, January 2005. ISSN 1529-2401. [122]

L. M. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, July 1992. ISSN 02628856. [52]

D. B. Fogel. *Evolutionary computation: toward a new philosophy of machine intelligence*. IEEE Press, 1995. ISBN 0780310381. [18]

A. Frank and A. Asuncion. UCI Machine Learning Repository, 2010. URL `http://archive.ics.uci.edu/ml`. [85]

A. L. N. Fred and A. K. Jain. Robust data clustering. *Engineering*, 2:II–128–II–133, 2003. ISSN 10636919. [17]

W. J. Freeman. Models of the dynamics of neural populations. *Electroencephalography and clinical neurophysiology. Supplement*, 34(1):9–18, January 1978. ISSN 0424-8155. [122]

W. J. Freeman and W. Schneider. Changes in Spatial Patterns of Rabbit Olfactory EEG with Conditioning to Odors. *Psychophysiology*, 19(1):44–56, January 1982. ISSN 0048-5772. [108]

Y. Freund, R. E. Schapire, and M. Hill. Experiments with a New Boosting Algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 1–9, 1996. [21, 169]

J. Friedman, T. Hastie, and R. Tibshirani. Special Invited Paper. Additive Logistic Regression: A Statistical View of Boosting: Discussion. *The annals of statistics*, 28(2):337–374, 2000. [20, 21, 22, 169]

M. Friedman. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32(200):675, December 1937. ISSN 01621459. [58]

J. E. Friedrich and T. E. Acree. Gas Chromatography Olfactometry (GC/O) of Dairy Products. *International Dairy Journal*, 8(3):235–241, March 1998. ISSN 09586946. [112]

R. W. Friedrich and S. I. Korsching. Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging. *Neuron*, 18(5):737–52, May 1997. ISSN 0896-6273. [120, 126]

R. W. Friedrich and G. Laurent. Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science*, 291(5505):889–94, February 2001. ISSN 0036-8075. [129]

R. W. Friedrich, C. J. Habermann, and G. Laurent. Multiplexing using synchrony in the zebrafish olfactory bulb. *Nature neuroscience*, 7(8):862–71, August 2004. ISSN 1097-6256. [121]

P. Fries, P. R. Roelfsema, a. K. Engel, P. König, and W. Singer. Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23): 12699–704, November 1997. ISSN 0027-8424. [48]

P. Fries, J. H. Reynolds, a. E. Rorie, and R. Desimone. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science (New York, N.Y.)*, 291(5508): 1560–3, February 2001. ISSN 0036-8075. [48]

K. J. Friston. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual review of neuroscience*, 25:221–50, January 2002. ISSN 0147-006X. [129]

R. a. Fuentes, M. I. Aguilar, M. L. Aylwin, and P. E. Maldonado. Neuronal activity of mitral-tufted cells in awake rats during passive and active odorant stimulation. *Journal of neurophysiology*, 100(1):422–30, July 2008. ISSN 0022-3077. [108]

S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '99*, pages 63–72, New York, New York, USA, 1999. ACM Press. ISBN 1581131437. [15]

B. G. Galef and L.-A. Giraldeau. Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. *Animal behaviour*, 61(1):3–15, January 2001. ISSN 0003-3472. [95]

C. Galizia, S. Sachse, A. Rappert, and R. Menzel. The glomerular code for odor representation is species specific in the honeybee Apis mellifera. *nature neuroscience*, 2:473–478, 1999. [125]

C. G. Galizia. Functional Imaging. In M. D. Binder, N. Hirokawa, and U. Windhorst, editors, *Encyclopedia of Neuroscience*, pages 1649–1651. Springer Berlin Heidelberg, 2009. [120]

C. G. Galizia, K. Nägler, B. Hölldobler, and R. Menzel. Odour coding is bilaterally symmetrical in the antennal lobes of honeybees (Apis mellifera). *The European journal of neuroscience*, 10(9):2964–74, September 1998. ISSN 0953-816X. [109]

J. W. Gardner, J. a. Covington, S.-L. Tan, and T. C. Pearce. Towards an artificial olfactory mucosa for improved odour classification. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2083):1713–1728, July 2007. ISSN 1364-5021. [102]

J. Gascuel and C. Masson. A quantitative ultrastructural study of the honeybee antennal lobe. *Tissue and Cell*, 23(3):341–355, 1991. ISSN 00408166. [99]

W. S. Geisler and R. L. Diehl. Bayesian natural selection and the evolution of perceptual systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 357(1420):419–48, April 2002. ISSN 0962-8436. [164]

S. Gelstein, Y. Yeshurun, L. Rozenkrantz, S. Shushan, I. Frumin, Y. Roth, and N. Sobel. Human tears contain a chemosignal. *Science*, 331(6014):226–230, 2011. [95]

A. P. Georgopoulos, A. Schwartz, and R. Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, September 1986. ISSN 0036-8075. [123]

S. Ghosh, S. D. Larson, H. Hefzi, Z. Marnoy, T. Cutforth, K. Dokka, and K. K. Baldwin. Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature*, 472(7342):217–220, 2011. [110]

A. N. Gilbert. *What the nose knows: the science of scent in everyday life*. Crown Publishers, New York, New York, USA, 2008. ISBN 9781400082346. [97, 113, 116]

A. N. Gilbert and S. J. Firestein. Dollars and scents: commercial opportunities in olfaction and taste. *Nature neuroscience*, 5 Suppl:1043–5, November 2002. ISSN 1097-6256. [96]

D. Gill and T. C. Pearce. Wiring the Olfactory Bulb - Activity-dependent Models of Axonal Targeting in the Developing Olfactory Pathway. *Reviews in the Neurosciences*, 14(1-2): 63–72, January 2003. ISSN 0334-1763. [107]

G. Glusman, I. Yanai, I. Rubin, and D. Lancet. The complete human olfactory subgenome. *Genome research*, 11(5):685–702, May 2001. ISSN 1088-9051. [102]

J. A. Gottfried. Smell: central nervous processing. *Advances in oto-rhino-laryngology*, 63 (p 337):44–69, January 2006. ISSN 0065-3071. [92]

J. A. Gottfried. Central mechanisms of odour object perception. *Nature reviews. Neuroscience*, 11(SepTemBeR), August 2010. ISSN 1471-0048. [110, 111]

J. A. Gottfried, J. S. Winston, and R. J. Dolan. Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron*, 49(3):467–79, February 2006. ISSN 0896-6273. [111]

F. Grabenhorst, E. T. Rolls, C. Margot, M. a. a. P. da Silva, and M. I. Velazco. How pleasant and unpleasant stimuli combine in different brain regions: odor mixtures. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(49): 13532–40, December 2007. ISSN 1529-2401. [110]

A. Graf and S. Borer. Normalization in support vector machines. In *Pattern Recognition*, volume 2191/2001, pages 277–282. Springer, 2001. [56]

B. G. Green, D. Nachtigal, S. Hammond, and J. Lim. Enhancement of retronasal odors by taste. *Chemical senses*, 37(1):77–86, January 2012. ISSN 1464-3553. [95]

A. Grinvald and R. Hildesheim. VSDI: a new era in functional imaging of cortical dynamics. *Nature reviews. Neuroscience*, 5(11):874–85, November 2004. ISSN 1471-003X. [121]

N. Gupta and M. Stopfer. Insect olfactory coding and memory at multiple timescales. *Current opinion in neurobiology*, 21(5):768–73, October 2011. ISSN 1873-6882. [120]

K. M. Guthrie and C. M. Gall. Functional mapping of odor-activated neurons in the olfactory bulb. *Chemical senses*, 20(2):271–82, April 1995. ISSN 0379-864X. [134]

R. Gutierrez-Osuna. Pattern analysis for machine olfaction: a review. *IEEE Sensors Journal*, 2(3):189–202, June 2002a. ISSN 1530-437X. [96]

R. Gutierrez-Osuna. A self-organizing model of chemotopic convergence for olfactory coding. *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology*, 1(October): 236–237, 2002b. [145]

I. Guyon, editor. *Feature extraction: foundations and applications*. Studies in fuzziness. Springer Heidelberg, 2006. ISBN 3540354875. [51]

I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(7-8):1157–1182, 2003. ISSN 15324435. [24]

I. Guyon, S. Gunn, and A. Ben-Hur. Result analysis of the nips 2003 feature selection challenge. In *Advances in Neural Information Processing Systems*, volume 17, pages 545–552. MIT Press, 2004. [24]

L. B. Haberly. Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chemical senses*, 26(5):551–76, June 2001. ISSN 0379-864X. [110, 131]

L. B. Haberly and J. L. Price. The axonal projection patterns of the mitral and tufted cells of the olfactory bulb in the rat. *Brain research*, 129(1):152–7, June 1977. ISSN 0006-8993. [98]

L. B. Haberly and J. M. Bower. Olfactory cortex: model circuit for study of associative memory? *Trends in neurosciences*, 12(7):258–64, July 1989. ISSN 0166-2236. [101]

R. Haddad, R. Khan, Y. K. Takahashi, K. Mori, D. Harel, and N. Sobel. A metric for odorant comparison. *Nature methods*, 5(5):425–9, May 2008a. ISSN 1548-7105. [114, 118, 143, 153, 161]

R. Haddad, H. Lapid, D. Harel, and N. Sobel. Measuring smells. *Current opinion in neurobiology*, 18(4):438–44, August 2008b. ISSN 0959-4388. [112, 114, 125]

R. Haddad, T. Weiss, R. Khan, B. Nadler, N. Mandairon, M. Bensafi, E. Schneidman, and N. Sobel. Global features of neural activity in the olfactory system form a parallel code that predicts olfactory behavior and perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(27):9017–26, July 2010. ISSN 1529-2401. [123, 132, 160]

K. Haegler, R. Zernecke, A. M. Kleemann, J. Albrecht, O. Pollatos, H. Brückmann, and M. Wiesmann. No fear no risk! Human risk behavior is affected by chemosensory anxiety signals. *Neuropsychologia*, 48(13):3901–8, November 2010. ISSN 1873-3514. [95]

E. a. Hallem and J. R. Carlson. The odor coding system of Drosophila. *Trends in genetics : TIG*, 20(9):453–9, September 2004. ISSN 0168-9525. [99]

J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Elsevier Science Publishing Company, Inc., third edition, 2011. ISBN 9780123814791. [11]

J. A. Hartigan. Minimum mutation fits to a given tree. *Biometrics*, 29(1):53–65, 1973. ISSN 0006341X. [17]

J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, 293(5539):2425–30, September 2001. ISSN 0036-8075. [123]

W. Hays. Human pheromones: have they been demonstrated? *Behavioral Ecology and Sociobiology*, 54(2):89–97, 2003. [94]

A. Hierlemann and R. Gutierrez-Osuna. Higher-Order Chemical Sensing. *Chemical Reviews*, 108(2):563–613, February 2008. ISSN 0009-2665. [96]

J. G. Hildebrand and G. M. Shepherd. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annual review of neuroscience*, 20(1):595–631, January 1997. ISSN 0147-006X. [93]

C. C. Hilgetag, G. a. Burns, M. a. O'Neill, J. W. Scannell, and M. P. Young. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 355(1393):91–110, January 2000. ISSN 0962-8436. [129]

G. Hinton and J. L. McClelland. Distributed representations. In R. D. E., J. L. McClelland, and P. R. Group, editors, *Parallel distributed processing*, pages 77–109. MIT Press, Cambridge, MA, January 1986. [123]

S. L. Ho, B. A. Johnson, and M. I. Leon. Long hydrocarbon chains serve as unique molecular features recognized by ventral glomeruli of the rat olfactory bulb. *The Journal of comparative neurology*, 498(1):16–30, September 2006. ISSN 0021-9967. [129, 132]

B. Hold and M. Schleidt. The importance of human odour in non-verbal communication. *Zeitschrift für Tierpsychologie*, 43(3):225–38, March 1977. ISSN 0044-3573. [95]

J. Holland. Genetic algorithms. *Scientific American*, 267(July):66—-72, 1992. [18]

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–8, April 1982. ISSN 0027-8424. [46, 122]

J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10):3088–92, May 1984. ISSN 0027-8424. [46, 47]

J. J. Hopfield. Hopfield network. *Scholarpedia*, 2(5):1977, 2007. ISSN 1941-6016. [46]

D. E. Hornung and M. M. Mozell. Factors influencing the differential sorption of odorant molecules across the olfactory mucosa. *The Journal of general physiology*, 69(3):343 –61, March 1977. ISSN 0022-1295. [102]

J. D. Howard, J. Plailly, M. Grueschow, J.-D. Haynes, and J. A. Gottfried. Odor quality coding and categorization in human posterior piriform cortex. *Nature neuroscience*, 12 (7):932–8, July 2009. ISSN 1546-1726. [111, 119, 131]

D. Howes. *The Varieties of sensory experience: a sourcebook in the anthropology of the senses*. Anthropological horizons. University of Toronto Press, 1991. ISBN 9780802059024. [91]

C. Hsu, C. Chang, C. Lin, and Others. A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taiwan, 2003. URL `http://www.cs.manchester.ac.uk/pgt/2011/COMP61011/materials/practicalsvmguide.pdf`. [56]

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–54, January 1962. ISSN 0022-3751. [127]

G. Huesa, R. Anadón, and J. Yáñez. Olfactory projections in a chondrostean fish, Acipenser baeri: an experimental study. *The Journal of comparative neurology*, 428(1):145–58, December 2000. ISSN 0021-9967. [109]

T. Hummel and S. Nordin. Olfactory disorders and their consequences for quality of life. *Acta Oto-laryngologica*, 125(2):116–121, January 2005. ISSN 0001-6489. [93]

T. Hummel, T. Futschik, J. Frasnelli, and K.-B. Hüttenbrink. Effects of olfactory function, age, and gender on trigeminally mediated sensations: a study based on the lateralization of chemosensory stimuli. *Toxicology Letters*, 140-141:273–280, April 2003. ISSN 03784274. [92]

C. Humphries, E. Liebenthal, and J. R. Binder. Tonotopic organization of human auditory cortex. *NeuroImage*, 50(3):1202–11, April 2010. ISSN 1095-9572. [127]

T. Imai and H. Sakano. Roles of odorant receptors in projecting axons in the mouse olfactory system. *Current opinion in neurobiology*, 17(5):507–15, October 2007. ISSN 0959-4388. [105]

L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–506, January 2000. ISSN 0042-6989. [50]

A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*, volume 355 of *Prentice Hall Advanced Reference Series*. Prentice Hall, 1988. ISBN 013022278X. [17, 86]

A. T. James and A. J. P. Martin. Gas-liquid partition chromatography; the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. *The Biochemical journal*, 50(5):679–90, March 1952. ISSN 0264-6021. [112]

R. Jenssen, K. Hild, D. Erdogmus, J. Principe, and T. Eltoft. Clustering using Renyi's entropy. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 1, pages 523–528. IEEE, 2003. ISBN 0-7803-7898-9. [16]

T. Joachims. Making large-Scale SVM Learning Practical. *Advances in Kernel Methods Support Vector Learning*, pages 169–184, 1999. [140]

B. a. Johnson and M. Leon. Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. *The Journal of comparative neurology*, 422(4):496–509, July 2000a. ISSN 0021-9967. [124, 125, 126, 142, 154]

B. a. Johnson and M. Leon. Odorant molecular length: one aspect of the olfactory code. *The Journal of comparative neurology*, 426(2):330–8, October 2000b. ISSN 0021-9967. [124, 126, 129]

B. A. Johnson and M. Leon. Odorant molecular length: one aspect of the olfactory code. *The Journal of Comparative Neurology*, 426(2):330–338, 2000c. [131, 162, 163]

B. A. Johnson and M. Leon. Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. *The Journal of comparative neurology*, 422(4):496–509, July 2000d. ISSN 0021-9967. [102, 114, 131, 162]

B. A. Johnson and M. I. Leon. Chemotopic odorant coding in a mammalian olfactory system. *The Journal of comparative neurology*, 503(1):1–34, July 2007. ISSN 0021-9967. [128, 131, 134, 151, 157, 159, 160]

B. A. Johnson, C. C. Woo, H. Duong, V. Nguyen, and M. I. Leon. A learned odor evokes an enhanced Fos-like glomerular response in the olfactory bulb of young rats. *Brain research*, 699(2):192–200, November 1995. ISSN 0006-8993. [126]

B. A. Johnson, C. C. Woo, and M. I. Leon. Spatial coding of odorant features in the glomerular layer of the rat olfactory bulb. *The Journal of comparative neurology*, 393(4): 457–71, April 1998. ISSN 0021-9967. [124, 125, 126, 142]

B. A. Johnson, C. C. Woo, E. E. Hingco, K. L. Pham, and M. I. Leon. Multidimensional chemotopic responses to n-aliphatic acid odorants in the rat olfactory bulb. *The Journal of comparative neurology*, 409(4):529–48, July 1999. ISSN 0021-9967. [123, 124, 125, 129, 130, 133, 134, 142]

B. A. Johnson, S. L. Ho, Z. Xu, J. S. Yihan, S. Yip, E. E. Hingco, and M. I. Leon. Functional mapping of the rat olfactory bulb using diverse odorants reveals modular responses to functional groups and hydrocarbon structural features. *The Journal of comparative neurology*, 449(2):180–94, July 2002. ISSN 0021-9967. [123, 124, 142]

B. A. Johnson, H. Farahbod, Z. Xu, S. Saber, and M. I. Leon. Local and global chemotopic organization: general features of the glomerular representations of aliphatic odorants differing in carbon number. *The Journal of comparative neurology*, 480(2):234–249, 2004. [123, 124, 129, 130, 131, 142]

B. A. Johnson, H. Farahbod, and M. I. Leon. Interactions between odorant functional group and hydrocarbon structure influence activity in glomerular response modules in the rat

olfactory bulb. *The Journal of comparative neurology*, 483(2):205–16, March 2005a. ISSN 0021-9967. [113, 124]

B. A. Johnson, H. Farahbod, S. Saber, and M. I. Leon. Effects of functional group position on spatial representations of aliphatic odorants in the rat olfactory bulb. *The Journal of comparative neurology*, 483(2):192–204, March 2005b. ISSN 0021-9967. [123, 124, 142]

B. A. Johnson, Z. Xu, P. Pancoast, J. Kwok, J. Ong, and M. I. Leon. Differential specificity in the glomerular response profiles for alicyclic, bicyclic, and heterocyclic odorants. *The Journal of comparative neurology*, 499(1):1–16, November 2006. ISSN 0021-9967. [133, 134]

B. A. Johnson, S. Arguello, and M. I. Leon. Odorants with multiple oxygen-containing functional groups and other odorants with high water solubility preferentially activate posterior olfactory bulb glomeruli. *The Journal of comparative neurology*, 502(3):468–82, May 2007. ISSN 0021-9967. [156]

B. A. Johnson, J. Ong, and M. I. Leon. Glomerular activity patterns evoked by natural odor objects in the rat olfactory bulb are related to patterns evoked by major odorant components. *The Journal of comparative neurology*, 518(9):1542–55, May 2010. ISSN 1096-9861. [126]

P. Joussain, A. Chakirian, F. Kermen, C. Rouby, and M. Bensafi. Physicochemical influence on odor hedonics: Where does it occur first? *Communicative & integrative biology*, 4(5): 563–5, September 2011. ISSN 1942-0889. [114]

J. Kaas. Topographic maps are fundamental to sensory processing. *Brain research bulletin*, 44(2):107–12, January 1997. ISSN 0361-9230. [128, 129]

J. Kaas. Why is Brain Size so Important: Design Problems and Solutions as Neocortex Gets Biggeror Smaller. *Brain and Mind*, 1(1):7–23, 2000. [92]

V. a. Kalatsky and M. P. Stryker. New paradigm for optical imaging: temporally encoded maps of intrinsic signal. *Neuron*, 38(4):529–45, May 2003. ISSN 0896-6273. [128]

P. Karlson and M. Luscher. Pheromones': a new term for a class of biologically active substances. *Nature*, 183(4653):55–6, January 1959. ISSN 0028-0836. [94]

M. Kaschube, M. Schnabel, S. Löwel, D. M. Coppola, L. E. White, and F. Wolf. Universality in the evolution of orientation columns in the visual cortex. *Science (New York, N.Y.)*, 330(6007):1113–6, November 2010. ISSN 1095-9203. [128]

H. Kashiwadani, Y. F. Sasaki, N. Uchida, and K. Mori. Synchronized oscillatory discharges of mitral/tufted cells with different molecular receptive ranges in the rabbit olfactory bulb. *Journal of neurophysiology*, 82(4):1786–92, October 1999. ISSN 0022-3077. [121]

L. C. Katz and J. C. Crowley. Development of cortical circuits: lessons from ocular dominance columns. *Nature reviews. Neuroscience*, 3(1):34–42, January 2002. ISSN 1471-003X. [105]

J. S. Kauer and A. R. Cinelli. Are there structural and functional modules in the vertebrate olfactory bulb? *Microscopy research and technique*, 24(2):157–67, February 1993. ISSN 1059-910X. [102, 115]

U. B. Kaupp. Olfactory signalling in vertebrates and insects: differences and commonalities. *Nature reviews. Neuroscience*, 11(3):188–200, March 2010. ISSN 1471-0048. [93, 112]

L. M. Kay and G. Laurent. Odor- and context-dependent modulation of mitral cell activity in behaving rats. *Nature neuroscience*, 2(11):1003–9, November 1999. ISSN 1097-6256. [108]

A. Keller and L. B. Vosshall. Better smelling through genetics: mammalian odor perception. *Current opinion in neurobiology*, 18(4):364–9, August 2008. ISSN 0959-4388. [92, 100]

A. Keller, H. Zhuang, Q. Chi, L. B. Vosshall, and H. Matsunami. Genetic variation in a human odorant receptor alters odour perception. *Nature*, 449(7161):468–72, September 2007. ISSN 1476-4687. [100]

G. Kempermann, K. Fabel, D. Ehninger, H. Babu, P. Leal-Galicia, A. Garthe, and S. a. Wolf. Why and how physical activity promotes experience-induced brain plasticity. *Frontiers in neuroscience*, 4(December):189, January 2010. ISSN 1662-453X. [99]

F. Kermen, A. Chakirian, C. Sezille, P. Joussain, G. Le Goff, A. Ziessel, M. Chastrette, N. Mandairon, A. Didier, C. Rouby, and M. Bensafi. Molecular complexity determines the number of olfactory notes and the pleasantness of smells. *Scientific Reports*, 1:1–5, December 2011. ISSN 2045-2322. [114]

M. A. Kerr and L. Belluscio. Olfactory experience accelerates glomerular refinement in the mammalian olfactory bulb. *Nature neuroscience*, 9(4):484–6, April 2006. ISSN 1097-6256. [105]

R. P. Kesner, P. E. Gilbert, and G. V. Wallenstein. Testing neural network models of memory with behavioral experiments. *Current opinion in neurobiology*, 10(2):260–5, April 2000. ISSN 0959-4388. [122]

A. G. Khan, K. Parthasarathy, and U. S. Bhalla. Odor representations in the mammalian olfactory bulb. *Wiley interdisciplinary reviews. Systems biology and medicine*, 2(5):603–11, 2010. ISSN 1939-005X. [109, 125, 158, 162, 163]

M. Khan, E. Vaes, and P. Mombaerts. Regulation of the Probability of Mouse Odorant Receptor Gene Choice. *Cell*, 147(4):907–921, November 2011. ISSN 00928674. [102]

R. M. Khan, C.-h. Luk, A. Flinker, A. Aggarwal, H. Lapid, R. Haddad, and N. Sobel. Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(37):10015–23, September 2007. ISSN 1529-2401. [114, 116, 152]

R. Khan and N. Sobel. Neural processing at the speed of smell. *Neuron*, 44(5):744–747, 2004. [122]

D. H. Kim, M. E. Phillips, A. Y. Chang, H. K. Patel, K. T. Nguyen, and D. C. Willhite. Lateral Connectivity in the Olfactory Bulb is Sparse and Segregated. *Frontiers in neural circuits*, 5(April):5, January 2011. ISSN 1662-5110. [99, 107, 135]

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science (New York, N.Y.)*, 220(4598):671–80, May 1983. ISSN 0036-8075. [80]

T. Knijnenburg. Selecting relevant and non-relevant features in microarray classification applications, 2004. [25]

P. Knüsel, M. a. Carlsson, B. S. Hansson, T. C. Pearce, and P. F. M. J. Verschure. Time and space are complementary encoding dimensions in the moth antennal lobe. *Network (Bristol, England)*, 18(1):35–62, March 2007. ISSN 0954-898X. [122]

G. Kobal and B. Kettenmann. Olfactory functional imaging and physiology. *International Journal of Psychophysiology*, 36(2):157–163, 2000. [114]

K. Kobayakawa, R. Kobayakawa, H. Matsumoto, Y. Oka, T. Imai, M. Ikawa, M. Okabe, T. Ikeda, S. Itohara, T. Kikusui, K. Mori, and H. Sakano. Innate versus learned odour processing in the mouse olfactory bulb. *Nature*, 450(7169):503–8, November 2007. ISSN 1476-4687. [132]

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982. ISSN 0340-1200. [106]

T. Kosaka and K. Kosaka. "Interneurons" in the olfactory bulb revisited. *Neuroscience research*, 69(2):93–9, February 2011. ISSN 1872-8111. [102, 106]

S. Kotsiantis and I. Zaharakis. Supervised machine learning: A review of classification techniques. *applications in computer*, 31(3), 2007. [140]

A. A. Koulakov. In search of the structure of human olfactory space. *Frontiers in Systems Neuroscience*, 5(September):1–8, 2011. ISSN 16625137. [118]

G. Kovács, B. Gulyás, I. Savic, D. I. Perrett, R. E. Cornwell, A. C. Little, B. C. Jones, D. M. Burt, V. Gál, and Z. Vidnyánszky. Smelling human sex hormone-like compounds affects face gender judgment of men. *Neuroreport*, 15(8):1275–7, June 2004. ISSN 0959-4965. [95]

P. Kovesi. Image Features From Phase Congruency Image Features From Phase Congruency. *Image (Rochester, N.Y.)*, (June), 1999. [53]

P. Kovesi. Edges Are Not Just Steps. *Pattern Recognition*, (January):23–25, 2002. [53]

K. Krishna, N. Murty, and Others. Genetic K-means algorithm. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 29(3):433–439, 1999. [18]

J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, March 1964. ISSN 0033-3123. [144]

W. J. Krzanowski and Y. T. Lai. A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics*, 44(1):23–34, 1988. ISSN 0006341X. [17]

S. Kullback and R. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. [40]

T. O. Kvalseth. Entropy and correlation: Some comments. *Ieee Transactions On Systems Man And Cybernetics*, 17(3):517–519, 1987. ISSN 00189472. [17]

J. Laishram, D. Avossa, R. Shahapure, and V. Torre. Mechanical computation in neurons. *Developmental neurobiology*, 69(11):731–51, September 2009. ISSN 1932-846X. [105]

D. Lancet. Vertebrate olfactory reception. *Annual review of neuroscience*, 9:329–55, January 1986. ISSN 0147-006X. [156]

L. J. Land and G. M. Shepherd. Autoradiographic analysis of olfactory receptor projections in the rabbit. *Brain Research*, 70(3):506–510, April 1974. ISSN 00068993. [156]

A. Lansner. Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. *Trends in Neurosciences*, 32(3):178–186, 2009. [122]

A. Lansner, S. Benjaminsson, and C. Johansson. From ANN to Biomimetic Information Processing. In *Biologically Inspired Signal Processing for Chemical Sensing*, volume 188, pages 33–43. Springer, 2009. ISBN 9783642001758. [145, 163]

H. Lapid, S. Shushan, A. Plotkin, H. Voet, Y. Roth, T. Hummel, E. Schneidman, and N. Sobel. Neural activity at the human olfactory epithelium reflects olfactory perception. *Nature Neuroscience*, 14(11):1455–1461, September 2011. ISSN 1097-6256. [119, 131]

M. Larsson and J. Willander. Autobiographical odor memory. *Annals of the New York Academy of Sciences*, 1170(c):318–23, July 2009. ISSN 1749-6632. [94]

M. Laska and P. Teubner. Olfactory discrimination ability for homologous series of aliphatic alcohols and aldehydes. *Chemical senses*, 24(3):263–70, June 1999. ISSN 0379-864X. [114]

M. Laska, D. Genzel, and A. Wieser. The number of functional olfactory receptor genes and the relative size of olfactory brain structures are poor predictors of olfactory discrimination performance with enantiomers. *Chemical senses*, 30(2):171–5, February 2005. ISSN 0379-864X. [92]

M. Laszlo and S. Mukherjee. A genetic algorithm that exchanges neighboring centers for k-means clustering. *Pattern Recognition Letters*, 28(16):2359–2366, December 2007. ISSN 01678655. [19]

G. Laurent. Olfactory processing: maps, time and codes. *Current Opinion in Neurobiology*, 7(4):547–553, August 1997. ISSN 09594388. [119]

G. Laurent. Olfactory network dynamics and the coding of multidimensional signals. *Nature Reviews Neuroscience*, 3(November):884–895, 2002. [122]

H. Lawless. THE SENSE OF SMELL IN FOOD QUALITY AND SENSORY EVALUATION. *Journal of Food Quality*, 14(1):33–60, February 1991. ISSN 0146-9428. [95]

A. Le Guérer. Olfaction and cognition: A philosophical and psychoanalytic view. *Olfaction, taste, and cognition*, pages 3–15, 2002. [91, 115]

L. Lebart. Discrimination through the regularized nearest cluster method. In Y. Dodge and J. Whittaker, editors, *Computational Statistics*, volume 1, pages 103–118. Physica Verlag, Vienna, 1992. [83]

Y. LeCun and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages 97–104. IEEE, 2004. ISBN 0-7695-2158-4. [52]

J. E. Ledoux. Amygdala. *Scholarpedia*, 3(4):2698, 2008. ISSN 1941-6016. [93]

D. Lee, L. Itti, C. Koch, and J. Braun. Attention activates winner-take-all competition among visual filters. *nature neuroscience*, 2:375–381, 1999. [40]

J. Lee and M. Verleysen. *Nonlinear dimensionality reduction*. Springer Heidelberg, first edition, 2007. ISBN 978-0-387-39350-6. [145]

H. Lei, T. a. Christensen, and J. G. Hildebrand. Spatial and temporal organization of ensemble representations for different odor classes in the moth antennal lobe. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(49):11108–19, December 2004. ISSN 1529-2401. [119]

M. I. Leon and B. A. Johnson. Olfactory coding in the mammalian olfactory bulb. *Brain Research Reviews*, 42(1):23–32, April 2003. ISSN 01650173. [114, 124]

M. I. Leon and B. A. Johnson. Glomerular Activity Response Archive, 2006. URL `http://gara.bio.uci.edu/`. [104, 118, 127, 133, 134, 136, 160]

M. I. Leon and B. A. Johnson. Is there a space-time continuum in olfaction? *Cellular and molecular life sciences : CMLS*, 66(13):2135–50, July 2009. ISSN 1420-9071. [119, 125, 162]

S. Lewandowsky, K. Oberauer, and G. D. a. Brown. No temporal decay in verbal short-term memory. *Trends in cognitive sciences*, 13(3):120–6, March 2009. ISSN 1364-6613. [94]

A. Li, L. Gong, and F. Xu. Brain-state-independent neural representation of peripheral stimulation in rat olfactory bulb. *Proceedings of the National Academy of Sciences of the United States of America*, 108(12):5087–92, March 2011. ISSN 1091-6490. [108, 158]

W. Li, I. Moallem, K. a. Paller, and J. A. Gottfried. Subliminal smells can guide social preferences. *Psychological science*, 18(12):1044–9, December 2007. ISSN 0956-7976. [95]

W. Li, J. D. Howard, T. B. Parrish, and J. A. Gottfried. Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. *Science (New York, N.Y.)*, 319(5871):1842–5, March 2008. ISSN 1095-9203. [114]

S. D. Liberles and L. B. Buck. A second class of chemosensory receptors in the olfactory epithelium. *Nature*, 442(7103):645–50, August 2006. ISSN 1476-4687. [95]

J. Liebelt and C. Schmid. Multi-view object class detection with a 3D geometric model. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1688–1695. IEEE, June 2010. ISBN 978-1-4244-6984-0. [52]

J. Liebelt, C. Schmid, and K. Schertler. Viewpoint-independent object class detection using 3D Feature Maps. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, June 2008. ISBN 978-1-4244-2242-5. [52]

D. Y. Lin, S. D. Shea, and L. C. Katz. Representation of natural stimuli in the rodent main olfactory bulb. *Neuron*, 50(6):937–49, June 2006. ISSN 0896-6273. [123, 126, 142]

J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. ISSN 00189448. [40]

O. Linde and T. Lindeberg. Composed Complex-Cue Histograms: An Investigation of the Information Content in Receptive Field Based Image Descriptors for Object Recognition. *Computer Vision and Image Understanding*, 116(4), December 2012. ISSN 10773142. [42]

T. Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1):225–270, 1994. ISSN 0266-4763. [51, 52, 53]

T. Lindeberg. Scale-space. In Benjamin Wah, editor, *Encyclopedia of Computer Science and Engineering*, volume IX, pages 2495–2504. Hoboken, New Jersey, 2008. ISBN 9780470050118. [51, 53]

T. Lindeberg. Generalized Gaussian Scale-Space Axiomatics Comprising Linear Scale-Space, Affine Scale-Space and Spatio-Temporal Scale-Space. *Journal of Mathematical Imaging and Vision*, 40(1):36–81, December 2010. ISSN 0924-9907. [125]

T. Lindeberg. Scale Invariant Feature Transform. *Scholarpedia*, 1(3):1300, 2012. [51, 53]

C. Linneus. Dissertatio medica de odores medicamentorum exhibens, 1752. URL `http://books.google.com/books/about/Dissertatio_medica_de_odores_medicamento.html?id=e44ZAAAAYAAJ`. [116]

C. Linster, B. A. Johnson, E. Yue, A. Morse, Z. Xu, E. E. Hingco, Y. Choi, M. Choi, A. Messiha, and M. I. Leon. Perceptual correlates of neural representations evoked by odorant enantiomers. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(24):9837–43, December 2001. ISSN 1529-2401. [124, 132]

C. Linster, B. A. Johnson, A. Morse, E. Yue, and M. I. Leon. Spontaneous versus reinforced olfactory discriminations. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(16):6842–5, August 2002. ISSN 1529-2401. [132]

H. Liu, H. Motoda, and L. Yu. Feature selection with selective sampling. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pages 395–402, San Francisco, CA, 2002. Morgan Kaufmann Publishers Inc. [77]

H.-c. Liu, J.-m. Yih, D.-b. Wu, and S.-w. Liu. Fuzzy possibility c-mean clustering algorithms based on complete mahalanobis distances. In *2008 International Conference on Wavelet Analysis and Pattern Recognition*, pages 50–55. IEEE, August 2008. ISBN 978-1-4244-2238-8. [15, 16]

P.-M. Lledo and S. Lagier. Adjusting neurophysiological computations in the adult olfactory bulb. *Seminars in cell developmental biology*, 17(4):443–453, 2006. [108]

M. Lloyd and K. Roen. 'When you smell smoke...': 'Risk factors' and fire safety in action. *Health, Risk & Society*, 4(2):139–153, July 2002. ISSN 1369-8575. [96]

S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982. ISSN 0018-9448. [13]

C. Lodovichi, L. Belluscio, and L. C. Katz. Functional topography of connections linking mirror-symmetric maps in the mouse olfactory bulb. *Neuron*, 38(2):265–76, 2003. ISSN 08966273. [109, 131]

M. Lopez-Sanchez, J. Cerquides, D. Masip, and A. Puig. Classification of biomedical high-resolution micro-ct images for direct volume rendering. In V. Devedzic, editor, *Artificial Intelligence and Applications*, pages 373–378. IASTED/ACTA Press, 2007. ISBN 978-0-88986-631-7. [52]

L. Louw, K. Roux, A. Tredoux, O. Tomic, T. Naes, H. H. Nieuwoudt, and P. van Rensburg. Characterization of selected South African young cultivar wines using FTMIR spectroscopy, gas chromatography, and multivariate data analysis. *Journal of agricultural and food chemistry*, 57(7):2623–32, April 2009. ISSN 1520-5118. [112]

Y. Lu, S. Lu, F. Fotouhi, Y. Deng, and S. J. Brown. FGKA : A Fast Genetic K-means Clustering Algorithm. In *Proceedings of the 2004 ACM symposium on Applied computing*, pages 1–2, 2004. [18]

S. X. Luo, R. Axel, and L. F. Abbott. Generating sparse and selective third-order responses in the olfactory system of the fly. *Proceedings of the National Academy of Sciences of the United States of America*, 107(23):10713–8, June 2010. ISSN 1091-6490. [107]

U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, August 2007. ISSN 0960-3174. [15]

J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium*, 233(233):281–297, 1967. [13]

P. Mahalanobis. On the generalized distance in statistics. In *Proceedings of the National Institute of Science, Calcutta*, volume 12, page 49, 1936. [83]

J. D. Mainland, E. A. Bremner, N. Young, B. N. Johnson, R. M. Khan, M. Bensafi, and N. Sobel. Olfactory plasticity: one nostril knows what the other learns. *Nature*, 419 (6909):802, October 2002. ISSN 0028-0836. [109]

R. Malach, I. Levy, and U. Hasson. The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4):176–184, April 2002. ISSN 13646613. [123, 127]

B. Malnic, J. Hirono, T. Sato, and L. B. Buck. Combinatorial Receptor Codes for Odors. *Cell*, 96(5):713–723, March 1999. ISSN 00928674. [100, 115]

C. V. D. Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100, December 1973. ISSN 0023-5946. [106]

F. M. Malvestuto. Statistical Treatment of the Information Content of a Database. *Information Systems Journal*, 11(3):211–223, 1986. [17]

A. M. Mamlouk and T. Martinetz. On the dimensions of the olfactory perception space. *Neurocomputing*, 58-60:1019–1025, June 2004. ISSN 09252312. [116]

O. Man, D. C. Willhite, C. J. Crasto, G. M. Shepherd, and Y. Gilad. A framework for exploring functional variability in olfactory receptor genes. *PloS one*, 2(8):e682, January 2007. ISSN 1932-6203. [118]

N. Mandairon, J. Poncelet, M. Bensafi, and A. Didier. Humans and mice express similar olfactory preferences. *PloS one*, 4(1):e4209, January 2009. ISSN 1932-6203. [116]

B. Mandelbrot. How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science (New York, N.Y.)*, 156(3775):636–8, May 1967. ISSN 0036-8075. [41]

R. L. D. Mántaras. A distance-based attribute selection measure for decision tree induction. *Machine Learning*, 6(1):81–92, 1991. ISSN 08856125. [17]

I. Manzini, C. Brase, T.-W. Chen, and D. Schild. Response profiles to amino acid odorants of olfactory glomeruli in larval Xenopus laevis. *The Journal of physiology*, 581(Pt 2): 567–79, June 2007. ISSN 0022-3751. [120]

A. Maresh, D. Rodriguez Gil, M. C. Whitman, and C. A. Greer. Principles of glomerular organization in the human olfactory bulb–implications for odor processing. *PloS one*, 3 (7):e2640, January 2008. ISSN 1932-6203. [92, 99]

T. W. Margrie and A. T. Schaefer. Theta oscillation coupled spike latencies yield computational vigour in a mammalian sensory system. *The Journal of physiology*, 546(Pt 2): 363–74, January 2003. ISSN 0022-3751. [121]

D. Marr. *Vision: a computational investigation into the human representation and processing of visual information*. Freeman, 1982. ISBN 9780716712671. [50]

W. H. Marshall, C. N. Woolsey, and P. Bard. Observations on cortical somatic sensory mechanisms of cat and monkey. *Journal of Neurophysiology*, 4(1):1–24, 1941. [127]

S. J. Martin, H. Helanterä, and F. P. Drijfhout. Is parasite pressure a driver of chemical cue diversity in ants? *Proceedings. Biological sciences / The Royal Society*, 278(1705): 496–503, February 2011. ISSN 1471-2954. [93]

M. Massler. Geriatric nutrition: The role of taste and smell in appetite. *The Journal of Prosthetic Dentistry*, 43(3):247–250, March 1980. ISSN 00223913. [95]

MathWorks. MATLAB kstest2 function, 2007a. URL `http://www.mathworks.se/help/toolbox/stats/kstest2.html`. [38]

MathWorks. Matlab friedman function, 2007b. URL `http://www.mathworks.se/help/toolbox/stats/friedman.html`. [61]

MathWorks. Matlab kstest2 function, 2007c. URL `http://www.mathworks.com/access/helpdesk/help/toolbox/stats/kstest2.html`. [38]

MathWorks. Matlab signrank function, 2007d. URL `http://www.mathworks.se/help/toolbox/stats/signrank.html`. [62]

MathWorks. Matlab signtest function, 2007e. URL `http://www.mathworks.se/help/toolbox/stats/signtest.html`. [41]

H. Matsumoto, K. Kobayakawa, R. Kobayakawa, T. Tashiro, K. Mori, H. Sakano, and K. Mori. Spatial arrangement of glomerular molecular-feature clusters in the odorant-receptor class domains of the mouse olfactory bulb. *Journal of Neurophysiology*, 103(6): 3490–3500, 2010. [124, 128, 131]

U. Maulik and S. Bandyopadhyay. Genetic algorithm-based clustering technique. *Pattern recognition*, 33(9):1455–1465, 2000. [18]

T. McLaughlin and D. D. M. O'Leary. Molecular gradients and development of retinotopic maps. *Annual review of neuroscience*, 28:327–55, January 2005. ISSN 0147-006X. [105]

M. C. Meilgaard. Prediction of flavor differences between beers from their chemical composition. *Journal of Agricultural and Food Chemistry*, 30(6):1009–1017, November 1982. ISSN 0021-8561. [116]

E. Meisami. A proposed relationship between increases in the number of olfactory receptor neurons, convergence ratio and sensitivity in the developing rat. *Brain research. Developmental brain research*, 46(1):9–19, March 1989. ISSN 0165-3806. [98, 99, 104]

E. Meisami and L. Safari. A quantitative study of the effects of early unilateral olfactory deprivation on the number and distribution of mitral and tufted cells and of glomeruli in the rat olfactory bulb. *Brain research*, 221(1):81–107, September 1981. ISSN 0006-8993. [99]

M. Meister and T. Bonhoeffer. Tuning and topography in an odor map on the rat olfactory bulb. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(4):1351–60, February 2001. ISSN 1529-2401. [115, 120, 131, 156, 157]

M. M. Merzenich. Orderly representation of the cochlea within primary auditory cortex in the cat. *The Journal of the Acoustical Society of America*, 55(S1):S86, 1974. ISSN 00014966. [127]

M. Migliore and G. M. Shepherd. Dendritic action potentials connect distributed dendrodendritic microcircuits. *Journal of computational neuroscience*, 24(2):207–21, April 2008. ISSN 0929-5313. [107]

P. Miklavc and T. Valentincic. Chemotopy of Amino Acids on the Olfactory Bulb Predicts Olfactory Discrimination Capabilities of Zebrafish Danio rerio. *Chemical senses*, pages 65–75, July 2011. ISSN 1464-3553. [126]

G.-l. Ming and H. Song. Adult neurogenesis in the mammalian central nervous system. *Annual review of neuroscience*, 28:223–50, January 2005. ISSN 0147-006X. [106]

G.-l. Ming, S. T. Wong, J. Henley, X.-b. Yuan, H.-j. Song, N. C. Spitzer, and M.-m. Poo. Adaptation in the chemotactic guidance of nerve growth cones. *Nature*, 417(6887):411–8, May 2002. ISSN 0028-0836. [105]

S. a. Mingoti and J. O. Lima. Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, 174(3):1742–1759, November 2006. ISSN 03772217. [14, 85]

K. Miyamichi, S. Serizawa, H. M. Kimura, and H. Sakano. Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(14):3586–92, April 2005. ISSN 1529-2401. [131]

K. Miyamichi, F. Amat, F. Moussavi, C. Wang, I. Wickersham, N. R. Wall, H. Taniguchi, B. Tasic, Z. J. Huang, Z. He, E. M. Callaway, M. a. Horowitz, and L. Luo. Cortical representations of olfactory input by trans-synaptic tracing. *Nature*, 472(7342):191–6, April 2011. ISSN 1476-4687. [110, 111]

N. Miyasaka, K. Morimoto, T. Tsubokawa, S.-i. Higashijima, H. Okamoto, and Y. Yoshihara. From the olfactory bulb to higher brain centers: genetic visualization of secondary olfactory pathways in zebrafish. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(15):4756–67, April 2009. ISSN 1529-2401. [110]

P. Mombaerts. Odorant receptor gene choice in olfactory sensory neurons: the one receptor-one neuron hypothesis revisited. *Current opinion in neurobiology*, 14(1):31–6, February 2004. ISSN 0959-4388. [98, 102]

P. Mombaerts. Axonal wiring in the mouse olfactory system. *Annual review of cell and developmental biology*, 22:713–37, January 2006. ISSN 1081-0706. [105]

A. Montagnini and A. Treves. The evolution of mammalian cortex, from lamination to arealization. *Brain Research Bulletin*, 60(4):387–393, March 2003. ISSN 03619230. [120]

L. Moreaux and G. Laurent. A simple method to reconstruct firing rates from dendritic calcium signals. *Frontiers in neuroscience*, 2(2):176–85, December 2008. ISSN 1662-453X. [135]

K. Mori. Relation of chemical structure to specificity of response in olfactory glomeruli. *Current opinion in neurobiology*, 5(4):467–74, August 1995. ISSN 0959-4388. [100]

K. Mori. The Olfactory Bulb: Coding and Processing of Odor Molecule Information. *Science*, 286(5440):711–715, October 1999. ISSN 00368075. [102, 108]

K. Mori. Grouping of odorant receptors: odour maps in the mammalian olfactory bulb. *Biochemical Society transactions*, 31(Pt 1):134–6, February 2003. ISSN 0300-5127. [124]

K. Mori and H. Sakano. How is the olfactory map formed and interpreted in the mammalian brain? *Annual review of neuroscience*, 34:467–99, January 2011. ISSN 1545-4126. [105, 130]

K. Mori and G. M. Shepherd. Emerging principles of molecular signal processing by mitral/tufted cells in the olfactory bulb. *Seminars in Cell Biology*, 5(1):65–74, February 1994. ISSN 10434682. [100, 125, 129, 131]

K. Mori, Y. K. Takahashi, K. M. Igarashi, and M. Yamaguchi. Maps of odorant molecular features in the Mammalian olfactory bulb. *Physiological reviews*, 86(2):409–33, April 2006a. ISSN 0031-9333. [128]

K. Mori, Y. K. Takahashi, K. M. Igarashi, and M. Yamaguchi. Maps of odorant molecular features in the Mammalian olfactory bulb. *Physiological reviews*, 86(2):409–33, April 2006b. ISSN 0031-9333. [114, 131, 156]

K. Mori, H. Matsumoto, Y. Tsuno, and K. M. Igarashi. Dendrodendritic synapses and functional compartmentalization in the olfactory bulb. *Annals of the New York Academy of Sciences*, 1170:255–8, July 2009. ISSN 1749-6632. [107, 131]

M. Morrone and R. Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6(5):303–313, December 1987. ISSN 01678655. [53]

M. P. Moshkin, N. a. Litvinova, a. V. Bedareva, M. S. Bedarev, E. a. Litvinova, and L. a. Gerlinskaya. Odor as an element of subjective assessment of attractiveness of young males and females. *Journal of Evolutionary Biochemistry and Physiology*, 47(1):69–82, April 2011. ISSN 0022-0930. [95]

M. M. Mozell, B. P. Smith, P. E. Smith, R. L. Sullivan, and P. Swender. Nasal chemoreception in flavor identification. *Archives of otolaryngology (Chicago, Ill. : 1960)*, 90(3): 367–73, September 1969. ISSN 0003-9977. [95]

M. M. Mozell, P. R. Sheehe, D. E. Hornung, P. F. Kent, S. L. Youngentob, and S. J. Murphy. "Imposed" and "inherent" mucosal activity patterns. Their composite representation of olfactory stimuli. *The Journal of general physiology*, 90(5):625–50, November 1987. ISSN 0022-1295. [102]

M. M. Mozell. Evidence for Sorption as a Mechanism of the Olfactory Analysis of Vapours. *Nature*, 203(4950):1181–1182, September 1964. ISSN 0028-0836. [102]

M. C. Mozer. Attractor Networks. In P. Wilken, T. Bayne, and A. Cleeremans, editors, *Companion to Consciousness*, pages 86–89. Oxford University Press Oxford, illustrate edition, 2009. ISBN 0198569513. [122]

L. R. Mujica-Parodi, H. H. Strey, B. Frederick, R. Savoy, D. Cox, Y. Botanov, D. Tolkunov, D. Rubin, and J. Weber. Chemosensory Cues to Conspecific Emotional Stress Activate Amygdala in Humans. *PLoS ONE*, 4(7):14, 2009. [95]

V. N. Murthy. Olfactory maps in the brain. *Annual review of neuroscience*, 34:233–58, January 2011. ISSN 1545-4126. [125]

S. Nagayama, Y. K. Takahashi, Y. Yoshihara, and K. Mori. Mitral and tufted cells differ in the decoding manner of odor maps in the rat olfactory bulb. *Journal of neurophysiology*, 91(6):2532–40, June 2004. ISSN 0022-3077. [108]

S. Nagayama, A. Enerva, M. L. Fletcher, A. V. Masurkar, K. M. Igarashi, K. Mori, and W. R. Chen. Differential axonal projection of mitral and tufted cells in the mouse main olfactory system. *Frontiers in neural circuits*, 4(September):1–8, January 2010. ISSN 1662-5110. [108]

H. Nagle, R. Gutierrez-Osuna, and S. Schiffman. The how and why of electronic noses. *Spectrum, IEEE*, 35(9):22–31, 1998. [96]

M. Najac, D. De Saint Jan, L. Reguero, P. Grandes, and S. Charpak. Monosynaptic and polysynaptic feed-forward inputs to mitral cells from olfactory sensory neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(24): 8722–9, June 2011. ISSN 1529-2401. [108, 135]

T. Nakashima, M. Akamatsu, A. Hatanaka, and T. Kiyohara. Attenuation of stress-induced elevations in plasma ACTH level and body temperature in rats by green odor. *Physiology & Behavior*, 80(4):481–488, January 2004. ISSN 00319384. [95]

K. Nara, L. R. Saraiva, X. Ye, and L. B. Buck. A Large-Scale Analysis of Odor Coding in the Olfactory Epithelium. *Journal of Neuroscience*, 31(25):9179–9191, June 2011. ISSN 0270-6474. [100]

J. C. Nawroth, C. A. Greer, W. R. Chen, S. B. Laughlin, and G. M. Shepherd. An energy budget for the olfactory glomerulus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(36):9790–800, September 2007. ISSN 1529-2401. [134]

M. Nei, Y. Niimura, and M. Nozawa. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature reviews. Genetics*, 9(12):951–63, December 2008. ISSN 1471-0064. [164]

M. Q. Nguyen, C. a. Marks, L. Belluscio, and N. J. P. Ryba. Early expression of odorant receptors distorts the olfactory circuitry. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(27):9271–9, July 2010. ISSN 1529-2401. [102]

H. Nienborg, H. Bridge, A. J. Parker, and B. G. Cumming. Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(9):2065–76, March 2004. ISSN 1529-2401. [158]

J. Niessing and R. W. Friedrich. Olfactory pattern classification by discrete neuronal network states. *Nature*, 465(7294):47–52, 2010. [119, 122, 126, 129]

A. Noble, R. Arnold, and B. Masuda. Progress towards a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, 35(2):4–6, 1984. [116]

G. E. Noether. *Introduction to Statistics — The Nonparametric Way*. Springer-Verlag New York Inc, 1991. [30, 41]

H. Nyquist. Certain Factors Affecting Telegraph Speed. *Transactions of the American Institute of Electrical Engineers*, XLIII:412–422, January 1924. ISSN 0096-3860. [42]

K. Ohki, S. Chung, P. Kara, M. Hübener, T. Bonhoeffer, and R. C. Reid. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, 442(7105):925–8, August 2006. ISSN 1476-4687. [127]

G. Ohloff. Chemistry of odor stimuli. *Experientia*, 42(3):271–9, March 1986. ISSN 0014-4754. [98, 113]

J. K. Olofsson, M. Rönnlund, S. Nordin, L. Nyberg, L.-G. Nilsson, and M. Larsson. Odor identification deficit as a predictor of five-year global cognitive change: interactive effects with age and ApoE-epsilon4. *Behavior genetics*, 39(5):496–503, September 2009. ISSN 1573-3297. [96]

S. R. Olsen and R. I. R. Wilson. Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature*, 452(7190):956–60, April 2008. ISSN 1476-4687. [107]

B. a. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Current opinion in neurobiology*, 14(4):481–7, August 2004. ISSN 0959-4388. [129]

L. Ott and M. Longnecker. *An introduction to statistical methods and data analysis*. Enhanced Web Assign Series. Cengage Learning, 6, illustr edition, 2008. ISBN 0495017582. [62, 138]

J. Pager, I. Giachetti, A. Holley, and J. Le Magnen. A selective control of olfactory bulb electrical activity in relation to food deprivation and satiety in rats. *Physiology & Behavior*, 9(4):573–579, October 1972. ISSN 00319384. [108]

F. Pain, B. L'heureux, and H. Gurden. Visualizing odor representation in the brain: a review of imaging techniques for the mapping of sensory activity in the olfactory glomeruli. *Cellular and molecular life sciences : CMLS*, May 2011. ISSN 1420-9071. [120]

C. Pantofaru, G. Dorko, C. Schmid, and M. Hebert. Combining Regions and Patches for Object Class Localization. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 23–23. IEEE, 2006. ISBN 0-7695-2646-2. [52]

M. Papadopoulou, S. Cassenaer, T. Nowotny, and G. Laurent. Normalization for sparse encoding of odors by a wide-field interneuron. *Science (New York, N.Y.)*, 332(6030): 721–5, May 2011. ISSN 1095-9203. [107]

D. J. Parkhurst and E. Niebur. Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, 19(3):783–789, February 2004. ISSN 0953-816X. [54]

T. R. Payne and P. Edwards. Implicit Feature Selection with the Value Difference Metric. In *Proceedings of the 13th European Conference on Artificial Intelligence, ECAI-98*, pages 450—-454. JOHN WILEY \& SONS LTD, 1998. [32]

T. C. Pearce. Computational parallels between the biological olfactory pathway and its analogue 'the electronic nose': Part II. Sensor-based machine olfaction. *Bio Systems*, 41 (2):69–90, January 1997a. ISSN 0303-2647. [96]

T. C. Pearce. Computational parallels between the biological olfactory pathway and its analogue 'The Electronic Nose': Part I. Biological olfaction. *Biosystems*, 41(1):43–67, January 1997b. ISSN 03032647. [96]

T. C. Pearce, P. Verschure, J. White, and J. Kauer. Robust stimulus encoding in olfactory processing: hyperacuity and efficient signal transmission. *Emergent neural computational architectures based on neuroscience*, pages 461–479, 2001. [98, 104]

H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE transactions on pattern analysis and machine intelligence*, 27(8):1226–38, August 2005. ISSN 0162-8828. [25, 43, 44, 77, 169]

P. Perona and J. Malik. Detecting and localizing edges composed of steps, peaks and roofs. In *[1990] Proceedings Third International Conference on Computer Vision*, pages 52–57. IEEE Comput. Soc. Press, 1990. ISBN 0-8186-2057-9. [53]

X. Perrotton, M. Sturzel, and M. Roux. Implicit hierarchical boosting for multi-view object detection. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 958–965. IEEE, June 2010. ISBN 978-1-4244-6984-0. [52]

K. Persaud and G. Dodd. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature*, 299(5881):352–355, September 1982. ISSN 0028-0836. [96]

L. Pessoa and R. Adolphs. Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature reviews. Neuroscience*, 11(11): 773–83, November 2010. ISSN 1471-0048. [93]

D. Pfitzner, R. Leibbrandt, and D. Powers. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19(3):361–394, July 2009. ISSN 0219-1377. [16]

a. J. Pinching and T. P. Powell. The neuropil of the glomeruli of the olfactory bulb. *Journal of cell science*, 9(2):347–77, September 1971. ISSN 0021-9533. [102]

J. Plailly, J. D. Howard, D. R. Gitelman, and J. A. Gottfried. Attention to odor modulates thalamocortical connectivity in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(20):5257–67, May 2008. ISSN 1529-2401. [110]

R. H. Porter, R. D. Balogh, J. M. Cernoch, and C. Franchi. Recognition of kin through characteristic body odors. *Chemical Senses*, 11(3):389–395, 1986. ISSN 0379-864X. [95]

J. L. Price. An autoradiographic study of complementary laminar patterns of termination of afferent fibers to the olfactory cortex. *The Journal of comparative neurology*, 150(1): 87–108, July 1973. ISSN 0021-9967. [109]

J. L. Price and T. P. Powell. The morphology of the granule cells of the olfactory bulb. *Journal of cell science*, 7(1):91–123, July 1970. ISSN 0021-9533. [107]

F. Pulvermüller. Brain mechanisms linking language and action. *Nature reviews. Neuroscience*, 6(7):576–82, July 2005. ISSN 1471-003X. [129]

D. Purves, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and S. M. Williams. *Neuroscience*. Sinauer Associates, Sunderland, MA, 2004. ISBN 9780878936953. [99, 128]

M. Rabinovich, R. Huerta, and G. Laurent. Neuroscience. Transient dynamics for neural processing. *Science (New York, N.Y.)*, 321(5885):48–50, July 2008. ISSN 1095-9203. [122]

Y. Rabinovich and A. Wigderson. Techniques for bounding the convergence rate of genetic algorithms. *Random Structures and Algorithms*, 14(2):111–138, March 1999. ISSN 1042-9832. [18]

C. Rajski. A metric space of discrete probability distributions. *Information and Control*, 4 (4):371–377, December 1961. ISSN 00199958. [17]

B. Raman and R. Gutierrez-Osuna. Relating Sensor Responses of Odorants to Their Organoleptic Properties by Means of a Biologically-Inspired Model of Receptor Neuron Convergence onto Olfactory Bulb. In A. Gutiérrez and S. Marco, editors, *Biologically Inspired Signal Processing for Chemical Sensing*, pages 93–108. Springer, Berlin, Heidelberg, 2009. [132]

W. M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971. ISSN 01621459. [17]

G. Rätsch, T. Onoda, and K.-R. Müller. Soft Margins for AdaBoost. *Machine Learning*, 42(3):287–320, 2001. [21]

N. Ravel, P. Chabaud, C. Martin, V. Gaveau, E. Hugues, C. Tallon-Baudry, O. Bertrand, and R. Gervais. Olfactory learning modifies the expression of odour-induced oscillatory responses in the gamma (60-90 Hz) and beta (15-40 Hz) bands in the rat olfactory bulb. *European Journal of Neuroscience*, 17(2):350–358, January 2003. ISSN 0953816X. [121]

E. S. Raymond. The Art of Unix Programming. *System*, 59(September):560, 2003. [62]

R. a. Reale and T. J. Imig. Tonotopic organization in auditory cortex of the cat. *The Journal of comparative neurology*, 192(2):265–91, July 1980. ISSN 0021-9967. [127]

P. Reinagel and a. M. Zador. Natural scene statistics at the centre of gaze. *Network (Bristol, England)*, 10(4):341–50, November 1999. ISSN 0954-898X. [54]

G. Reineccius and Z. M. Vickers. Flavor characterization. In *Kirk-Othmer Encyclopedia of Chemical Technology*, volume 18, pages 25–25. John Wiley & Sons, Inc., 2000. ISBN 9780471238966. [112]

K. J. Ressler, S. L. Sullivan, and L. B. Buck. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell*, 79(7):1245–1255, 1994. [102]

M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *IEEE International Conference on Neural Networks*, 1993(3): 586–591, 1993. [47]

H. M. Robertson and K. W. Wanner. The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family. *Genome research*, 16(11):1395–403, November 2006. ISSN 1088-9051. [99]

C. Rochefort, G. Gheusi, J.-D. Vincent, and P.-M. Lledo. Enriched odor exposure increases the number of newborn neurons in the adult olfactory bulb and improves odor memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(7): 2679–89, April 2002. ISSN 1529-2401. [158]

F. Röck, N. Barsan, and U. Weimar. Electronic Nose:Âă Current Status and Future Trends. *Chemical Reviews*, 108(2):705–725, February 2008. ISSN 0009-2665. [96]

I. Rodriguez, C. A. Greer, M. Y. Mok, and P. Mombaerts. A putative pheromone receptor gene expressed in human olfactory mucosa. *Nature genetics*, 26(1):18–9, September 2000. ISSN 1061-4036. [95]

E. T. Rolls. Chemosensory Learning in the Cortex. *Frontiers in Systems Neuroscience*, 5 (September):1–13, 2011. ISSN 1662-5137. [110]

G. Romani, S. Williamson, and L. Kaufman. Tonotopic organization of the human auditory cortex. *Science*, 216(4552):1339–1340, June 1982. ISSN 0036-8075. [127]

K. J. Rossiter. StructureâĹŠOdor Relationships. *Chemical Reviews*, 96(8):3201–3240, January 1996. ISSN 0009-2665. [113]

P. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(1):53–65, 1987. ISSN 03770427. [17]

J. P. Royet, C. Souchier, F. Jourdan, and H. Ploye. Morphometric study of the glomerular population in the mouse olfactory bulb: numerical density and size distribution along the rostrocaudal axis. *The Journal of comparative neurology*, 270(4):559–68, April 1988. ISSN 0021-9967. [99]

J. P. Royet, H. Distel, R. Hudson, and R. Gervais. A re-estimation of the number of glomeruli and mitral cells in the olfactory bulb of rabbit. *Brain research*, 788(1-2):35–42, March 1998. ISSN 0006-8993. [99]

B. Rubin and L. Katz. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron*, 23(3):499–511, 1999. [109, 123, 124, 134, 142, 154, 156, 157]

D. B. Rubin and T. A. Cleland. Dynamical mechanisms of odor processing in olfactory bulb mitral cells. *Journal of neurophysiology*, 96(2):555–68, August 2006. ISSN 0022-3077. [119]

B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, 77 (1-3):157–173, 2007. ISSN 09205691. [56]

J. H. Ruth. Odor thresholds and irritation levels of several chemical substances: a review. *American Industrial Hygiene Association journal*, 47(3):A142–51, March 1986. ISSN 0002-8894. [96]

S. Sachse, a. Rappert, and C. G. Galizia. The spatial representation of chemical structures in the antennal lobe of honeybees: steps towards the olfactory code. *The European journal of neuroscience*, 11(11):3970–82, November 1999. ISSN 0953-816X. [99, 115, 120, 162, 163]

Y. Saeys, I. n. Inza, and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics Oxford England*, 23(19):2507–2517, 2007. [24, 25]

H. Sakano. Neural map formation in the mouse olfactory system. *Neuron*, 67(4):530–42, August 2010. ISSN 1097-4199. [105, 132]

E. Salinas and L. F. Abbott. Vector reconstruction from firing rates. *Journal of computational neuroscience*, 1(1-2):89–107, June 1994. ISSN 0929-5313. [124]

M. Sam, S. Vora, B. Malnic, W. Ma, M. V. Novotny, and L. B. Buck. Neuropharmacology. Odorants may arouse instinctive behaviours. *Nature*, 412(6843):142, July 2001. ISSN 0028-0836. [94]

G. Sanchez-Andrade and K. M. Kendrick. The main olfactory system and social learning in mammals. *Behavioural brain research*, 200(2):323–35, June 2009. ISSN 1872-7549. [93]

M. a. Sánchez-Montañés and T. C. Pearce. Why do olfactory neurons have unspecific receptive fields? *Bio Systems*, 67(1-3):229–38, 2002. ISSN 0303-2647. [100, 144, 164]

M. Sandström, A. Lansner, J. Hellgren-Kotaleski, and J.-P. Rospars. Modeling the response of a population of olfactory receptor neurons to an odorant. *Journal of Computational Neuroscience*, 27(3):337–355, 2009. [144]

J. R. Sanes and T. Jessell. The Guidance of Axons to Their Targets. In E. R. Kandel, J. H. Schwartz, and T. Jessell, editors, *Principles of Neural Science*, pages 1063–1086. McGraw-Hill, New York, NY, 2000. ISBN 0838577016. [105, 106]

S. Savarese. 3D generic object categorization, localization and pose estimation. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. ISBN 978-1-4244-1630-1. [52]

I. Savic and H. Berglund. Passive perception of odors and semantic circuits. *Human brain mapping*, 21(4):271–8, April 2004. ISSN 1065-9471. [110]

I. Savic and H. Berglund. Androstenol–a steroid derived odor activates the hypothalamus in women. *PloS one*, 5(2):e8651, January 2010. ISSN 1932-6203. [95]

I. Savic, B. Gulyas, M. Larsson, and P. Roland. Olfactory Functions Are Mediated by Parallel and Hierarchical Processing. *Neuron*, 26(3):735–745, June 2000. ISSN 08966273. [110]

I. Savic, H. Berglund, B. Gulyas, and P. Roland. Smelling of Odorous Sex Hormone-like Compounds Causes Sex-Differentiated Hypothalamic Activations in Humans. *Neuron*, 31(4):661–668, August 2001. ISSN 08966273. [93, 95]

B. Schaal, G. Coureaud, D. Langlois, C. Giniès, E. Sémon, and G. Perrier. Chemical and behavioural characterization of the rabbit mammary pheromone. *Nature*, 424(6944): 68–72, July 2003. ISSN 1476-4687. [94]

A. T. Schaefer and T. W. Margrie. Spatiotemporal representations in the olfactory system. *Trends in neurosciences*, 30(3):92–100, March 2007. ISSN 0166-2236. [121, 122]

R. E. Schapire. The Strength of Weak Learnability. *Machine Learning*, 5:197–227, 1990. [20]

R. E. Schapire, P. Avenue, and A. Room. The Boosting Approach to Machine Learning An Overview. In *Nonlinear Estimation and Classification*. 2001. [21]

P. Scheunders. A genetic c-Means clustering algorithm applied to color image quantization. *Pattern Recognition*, 30(6):859–866, June 1997. ISSN 00313203. [18]

C. E. Schleich and R. Zenuto. Testing detection and discrimination of vegetation chemical cues in the subterranean rodent Ctenomys talarum. *Ethology Ecology & Evolution*, 22 (3):257–264, August 2010. ISSN 0394-9370. [93]

A. Schloegl. spearman function for matlab, 2002. URL `http://www.koders.com/matlab/fid3BBEB656ADADF78E2EFF724423D83AE5699159FB.aspx`. [31]

M. Schmuker and G. Schneider. Processing and classification of chemical data inspired by insect olfaction. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51):20285–9, December 2007. ISSN 1091-6490. [144, 145]

T. A. Schoenfeld and T. K. Knott. Evidence for the disproportionate mapping of olfactory airspace onto the main olfactory bulb of the hamster. *The Journal of comparative neurology*, 476(2):186–201, 2004. [130, 131]

N. E. Schoppa and G. L. Westbrook. Glomerulus-specific synchronization of mitral cells in the olfactory bulb. *Neuron*, 31(4):639–51, August 2001. ISSN 0896-6273. [121]

N. E. Schoppa. Synchronization of olfactory bulb mitral cells by precisely timed inhibitory inputs. *Neuron*, 49(2):271–83, January 2006. ISSN 0896-6273. [121]

C. S. Sell. On the unpredictability of odor. *Angewandte Chemie (International ed. in English)*, 45(38):6254–61, September 2006. ISSN 1433-7851. [113, 125]

J. L. Semmelhack and J. W. Wang. Select Drosophila glomeruli mediate innate olfactory attraction and aversion. *Nature*, 459(7244):218–23, May 2009. ISSN 1476-4687. [132]

M. I. Sereno, S. Pitzalis, and a. Martinez. Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science (New York, N.Y.)*, 294(5545): 1350–4, November 2001. ISSN 0036-8075. [127]

M. I. Sereno. Cognitive Science Society . Language And The Primate Brain. In N. Hillsdale, editor, *Proceedings Thirteenth Annual Conference of the Cognitive Science Society*, pages 79–84. Erlbaum Associates, 1991. [128]

S. Serizawa, K. Miyamichi, and H. Sakano. One neuron-one receptor rule in the mouse olfactory system. *Trends in genetics : TIG*, 20(12):648–53, December 2004. ISSN 0168-9525. [98]

S. Serizawa, K. Miyamichi, H. Takeuchi, Y. Yamagishi, M. Suzuki, and H. Sakano. A neuronal identity code for the odorant receptor-specific and activity-dependent axon sorting. *Cell*, 127(5):1057–69, December 2006. ISSN 0092-8674. [105, 106]

R. H. Sheikh, M. Raghuwanshi, and A. N. Jaiswal. Genetic Algorithm Based Clustering: A Survey. *2008 First International Conference on Emerging Trends in Engineering and Technology*, 2(6):314–319, July 2008. [18]

G. M. Shepherd. Synaptic organization of the mammalian olfactory bulb. *Physiological Reviews*, 52(4):864, 1972. [156]

G. M. Shepherd. A molecular vocabulary for olfaction. *Annals of the New York Academy of Sciences*, 510:98–103, January 1987. ISSN 0077-8923. [100]

G. M. Shepherd. Neurobiology. Modules for molecules. *Nature*, 358(6386):457–8, August 1992. ISSN 0028-0836. [102]

G. M. Shepherd. The human sense of smell: are we better than we think? *PLoS biology*, 2 (5):E146, May 2004. ISSN 1545-7885. [91]

G. M. Shepherd, W. Chen, and C. A. Greer. The Olfactory Bulb. In *The Synaptic Organization of the Brain*, pages 165—-217. Press, Oxford University, 2004. [100, 103, 106, 107]

G. M. Shepherd, W. R. Chen, D. C. Willhite, M. Migliore, and C. A. Greer. The olfactory granule cell: from classical enigma to central role in olfactory processing. *Brain research reviews*, 55(2):373–82, October 2007. ISSN 0165-0173. [107]

M. Shipley, M. Ennis, and A. Puche. The olfactory system. In G. Paxinos, editor, *The Rat Nervous System (Third Edition)*, pages 611–622. Elsevier Inc., San Diego, CA, 2008. ISBN 978-0-12-547638-6. [101, 110]

W. K. Simmons and L. W. Barsalou. The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cognitive neuropsychology*, 20(3):451–86, May 2003. ISSN 0264-3294. [129]

W. Singer. The organization of sensory motor representations in the neocortex: A hypothesis based on temporal coding. In C. Umiltà and M. Moscovitch, editors, *Attention and performance XV: Conscious and nonconscious information processing*, pages 77–107. MIT Press Cambridge, MA, USA, 1994. ISBN 978-0262210126. [127]

W. Singer. Time as coding space ? *Current Opinion in Neurobiology*, 9(2):189–194, 1999. [122]

M. Sitzmann. Chemical Identifier Resolver, 2009. URL `http://cactus.nci.nih.gov/`. [113]

B. M. Slotnick, H. Panhuber, G. a. Bell, and D. G. Laing. Odor-induced metabolic activity in the olfactory bulb of rats trained to detect propionic acid vapor. *Brain research*, 500 (1-2):161–8, October 1989. ISSN 0006-8993. [134]

T. D. Smith and K. P. Bhatnagar. Microsmatic primates: reconsidering how and when size matters. *Anatomical record. Part B, New anatomist*, 279(1):24–31, July 2004. ISSN 1552-4906. [92]

T. D. Smith, J. B. Rossie, and K. P. Bhatnagar. Evolution of the nose and nasal skeleton in primates. *Evolutionary Anthropology: Issues, News, and Reviews*, 16(4):132–146, August 2007. ISSN 10601538. [92]

Z. Soh, T. Tsuji, N. Takiguchi, and H. Ohtake. An artificial neural network approach for glomerular activity pattern prediction using the graph kernel method and the gaussian mixture functions. *Chemical senses*, 36(5):413–24, June 2011. ISSN 1464-3553. [114]

D. L. Sosulski, M. L. Bloom, T. Cutforth, R. Axel, and S. R. Datta. Distinct representations of olfactory information in different cortical centres. *Nature*, 472(7342):213–6, April 2011. ISSN 1476-4687. [110]

E. R. Soucy, D. F. Albeanu, A. L. Fantana, V. N. Murthy, and M. Meister. Precision and diversity in an odor map on the olfactory bulb. *Nature Neuroscience*, 12(2):210–220, 2009. [109, 128, 134, 157]

E. Spangenberg and B. Grohmann. ItâĂŹs Beginning to Smell (and Sound) a Lot Like Christmas: The Interactive Effects of Ambient Scent and Music in a Retail Setting. *Journal of Business Research*, (December 2003):1–20, 2005. [97]

E. Spangenberg, D. Sprott, B. Grohmann, and D. Tracy. Gender-congruent ambient scent influences on approach and avoidance behaviors in a retail storeâŸĘ. *Journal of Business Research*, 59(12):1281–1287, November 2006. ISSN 01482963. [97]

M. Spehr. Sniffing out social signals. *e-Neuroforum*, 1(1):9–16, March 2010. ISSN 1868-856X. [94, 102]

R. W. Sperry. Chemoaffinity in the Orderly Growth of Nerve Fiber Patterns and Connections. *Proceedings of the National Academy of Sciences of the United States of America*, 50:703–10, October 1963. ISSN 0027-8424. [105, 145]

R. W. Sperry. Optic nerve regeneration with return of vision in anurans. *Journal of Neurophysiology*, 7(1):57–69, 1944. [105]

R. W. Sperry. Restoration of vision after crossing of optic nerves and after contralateral transplantation of eye. *Journal of Neurophysiology*, 8(1):15–28, 1945. [105]

O. Sporns. *Networks of the Brain*, volume 1. MIT Press, 2010. ISBN 9780262014694. [120, 129]

H. Spors and A. Grinvald. Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron*, 34(2):301–15, April 2002. ISSN 0896-6273. [115, 119, 120, 122, 126, 134]

C. Stanfill and D. Waltz. Toward memory-based reasoning. *Communications of the ACM*, 29(12):1213–1228, December 1986. ISSN 00010782. [29, 32]

K. Stern and M. K. McClintock. Regulation of ovulation by human pheromones. *Nature*, 392(6672):177–9, March 1998. ISSN 0028-0836. [95]

D. D. Stettler and R. Axel. Representations of odor in the piriform cortex. *Neuron*, 63(6): 854–64, September 2009. ISSN 1097-4199. [110]

W. B. Stewart, J. S. Kauer, and G. M. Shepherd. Functional organization of rat olfactory bulb analysed by the 2-deoxyglucose method. *The Journal of comparative neurology*, 185 (4):715–34, June 1979. ISSN 0021-9967. [120, 124, 125, 154]

R. F. Stocker. The organization of the chemosensory system in Drosophila melanogaster: a rewiew. *Cell and Tissue Research*, 275(1):3–26, January 1994. ISSN 0302-766X. [99]

P. Stockinger, D. Kvitsiani, S. Rotkopf, L. Tirián, and B. J. Dickson. Neural circuitry that governs Drosophila male courtship behavior. *Cell*, 121(5):795–807, June 2005. ISSN 0092-8674. [132]

L. Stowers and D. W. Logan. Olfactory mechanisms of stereotyped behavior: on the scent of specialized circuits. *Current opinion in neurobiology*, 20(3):274–80, June 2010. ISSN 1873-6882. [94]

N. J. Strausfeld and J. G. Hildebrand. Olfactory systems: common design, uncommon origins? *Current opinion in neurobiology*, 9(5):634–9, October 1999. ISSN 0959-4388. [93]

A. Strehl and J. Ghosh. Cluster EnsemblesA Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3(3):583–617, 2003. ISSN 15324435. [17]

J. Strotmann, S. Conzelmann, A. Beck, P. Feinstein, H. Breer, and P. Mombaerts. Local permutations in the glomerular array of the mouse olfactory bulb. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(18):6927–38, September 2000. ISSN 0270-6474. [98, 102]

N. V. Swindale. The development of topography in the visual cortex: a review of models. *Network (Bristol, England)*, 7(2):161–247, May 1996. ISSN 0954-898X. [106]

N. V. Swindale. Visual map. *Scholarpedia*, 3(6):4607, 2008. ISSN 1941-6016. [127]

Y. K. Takahashi, M. Kurosaki, S. Hirono, and K. Mori. Topographic representation of odorant molecular features in the rat olfactory bulb. *Journal of neurophysiology*, 92(4): 2413–27, October 2004. ISSN 0022-3077. [120, 124]

S. Talbot and W. Marshall. Physiological studies on neural mechanisms of visual localization and discrimination. *Am. J. Ophthalmol*, 24(1255-1264):389, 1941. [127]

J. Tan, A. Savigner, M. Ma, and M. Luo. Odor information processing by the olfactory bulb analyzed in gene-targeted mice. *Neuron*, 65(6):912–26, March 2010. ISSN 1097-4199. [108, 124]

D. Tank and J. J. Hopfield. Simple 'neural' optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Transactions on Circuits and Systems*, 33(5):533–541, May 1986. ISSN 0098-4094. [46]

J.-P. Thivierge and G. F. Marcus. The topographic brain: from neural connectivity to cognition. *Trends in neurosciences*, 30(6):251–9, June 2007. ISSN 0166-2236. [126, 127, 128, 129]

R. Tirindelli, M. Dibattista, S. Pifferi, and A. Menini. From pheromones to behavior. *Physiological reviews*, 89(3):921, 2009. [93]

J. Todrank, D. Byrnes, A. Wrzesniewski, and P. Rozin. Odors can change preferences for people in photographs: A cross-modal evaluative conditioning study with olfactory USs and visual CSs*1. *Learning and Motivation*, 26(2):116–140, May 1995. ISSN 00239690. [95]

R. B. Tootell, E. Switkes, M. S. Silverman, and S. L. Hamilton. Functional anatomy of macaque striate cortex. II. Retinotopic organization. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 8(5):1531–68, May 1988. ISSN 0270-6474. [127]

K. Torkkola. Feature Extraction by Non-Parametric Mutual Information Maximization. *Journal of Machine Learning Research*, 3(7-8):1415–1438, 2003. [30]

G. Toussaint. Note on optimal selection of independent binary-valued features for pattern recognition (Corresp.). *IEEE Transactions on Information Theory*, 17(5):618–618, September 1971. ISSN 0018-9448. [25, 28]

a. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, January 1980. ISSN 0010-0285. [50]

D. Trotier. Vomeronasal organ and human pheromones. *European annals of otorhinolaryngology, head and neck diseases*, 128(4):184–190, March 2011. ISSN 1879-7296. [95]

D. Tsigankov and A. a. Koulakov. Sperry versus Hebb: topographic mapping in Isl2/EphA3 mutant mice. *BMC neuroscience*, 11(1):155, January 2010. ISSN 1471-2202. [106]

L. Turin. A Spectroscopic Mechanism for Primary Olfactory Reception. *Chemical Senses*, 21(6):773–791, 1996. ISSN 0379-864X. [100]

L. Turin. A method for the calculation of odor character from molecular structure. *Journal of theoretical biology*, 216(3):367–85, June 2002. ISSN 0022-5193. [113]

L. Turin. *The secret of scent: adventures in perfume and the science of smell*. Ecco, illustrate edition, 2006. ISBN 0061133833. [113]

R. J. Tusa, L. a. Palmer, and a. C. Rosenquist. The retinotopic organization of area 17 (striate cortex) in the cat. *The Journal of comparative neurology*, 177(2):213–35, January 1978. ISSN 0021-9967. [127]

N. Uchida and Z. F. Mainen. Speed and accuracy of olfactory discrimination in the rat. *Nature neuroscience*, 6(11):1224–9, November 2003. ISSN 1097-6256. [122]

N. Uchida, Y. K. Takahashi, M. Tanifuji, and K. Mori. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nature neuroscience*, 3(10):1035–43, October 2000. ISSN 1097-6256. [115, 120, 124, 131]

S. B. Udin and J. W. Fawcett. Formation of topographic maps. *Annual review of neuroscience*, 11(1):289–327, January 1988. ISSN 0147-006X. [127]

D. Valentin, C. Chrea, D. Nguyen, and Others. Taste-odour interactions in sweet taste perception. *Optimising sweet taste in foods. Cambridge (UK): Woodhead Publishing Ltd*, (March 2006):66–84, 2006. [95]

O. Valenzuela, I. Rojas, L. J. Herrera, A. Guillén, F. Rojas, L. Marquez, and M. Pasadas. Feature Selection Using Mutual Information and Neural Networks. *Monografías del Seminario Matemático García de Galdeano*, 33:331–340, 2006. [45]

M. Valley. Mouse MOB stain TOTO-3, 2006. URL `http://commons.wikimedia.org/wiki/File:Mouse_MOB_three_color.jpg`. [103]

W. van der Goes van Naters and J. R. Carlson. Receptors and neurons for fly odors in Drosophila. *Current biology : CB*, 17(7):606–12, April 2007. ISSN 0960-9822. [132]

B. van Ginneken, B. M. ter Haar Romeny, and M. a. Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE transactions on medical imaging*, 20(12):1228–41, December 2001. ISSN 0278-0062. [51]

V. N. Vapnik. *The Nature of Statistical Learning Theory*, volume 8 of *Statistics for Engineering and Information Science*. Springer, 1995. ISBN 0387945598. [25, 28]

H. Varendi, R. Porter, and J. Winberg. Does the newborn baby find the nipple by smell? *The Lancet*, 344(8928):989–990, October 1994. ISSN 01406736. [95]

R. Varshavsky, A. Gottlieb, M. Linial, and D. Horn. Novel unsupervised feature filtering of biological data. *Bioinformatics (Oxford, England)*, 22(14):e507–13, July 2006. ISSN 1367-4811. [84]

R. Vassar, S. K. Chao, R. Sitcheran, J. M. Nuñez, L. B. Vosshall, and R. Axel. Topographic organization of sensory projections to the olfactory bulb. *Cell*, 79(6):981–91, December 1994. ISSN 0092-8674. [102, 109, 131, 156]

M. Vogt. LibSVM matlab interface, 2004. URL `http://pc228.rt.e-technik.tu-darmstadt.de/~vogt/de/`. [56]

L. B. Vosshall, a. M. Wong, and R. Axel. An olfactory sensory map in the fly brain. *Cell*, 102(2):147–59, July 2000. ISSN 0092-8674. [99]

V. S. Vyas and P. Rege. Automated Texture Analysis with Gabor filter. *ICGST International Journal on Graphics, Vision and Image Processing*, 6(1):35–41, 2006. [51]

M. Wachowiak and L. B. Cohen. Representation of Odorants by Receptor Neuron Input to the Mouse Olfactory Bulb. *Neuron*, 32(4):723–735, November 2001. ISSN 08966273. [115, 120]

M. Wachowiak and L. B. Cohen. Correspondence between odorant-evoked patterns of receptor neuron input and intrinsic optical signals in the mouse olfactory bulb. *Journal of neurophysiology*, 89(3):1623–39, March 2003. ISSN 0022-3077. [115, 120]

M. P. Wand. Data-Based Choice of Histogram Bin Width. *The American Statistician*, 51 (1):59, February 1997. ISSN 00031305. [42, 137]

B. a. Wandell, A. a. Brewer, and R. F. Dougherty. Visual field map clusters in human cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1456):693–707, April 2005. ISSN 0962-8436. [127]

N. Wang. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90(1):43–52, March 2003. ISSN 0006-3444. [15]

Z. Wang, A. Nudelman, and D. R. Storm. Are Pheromones Detected Through the Main Olfactory Epithelium? *Molecular Neurobiology*, 35(3):317–323, June 2007. ISSN 0893-7648. [95]

Z. S. Warwick, W. G. Hall, T. N. Pappas, and S. S. Schiffman. Taste and smell sensations enhance the satiating effect of both a high-carbohydrate and a high-fat meal in humans. *Physiology & behavior*, 53(3):553–63, March 1993. ISSN 0031-9384. [95]

C. Watson, M. Kirkcaldie, and G. Paxinos. *The Brain: An Introduction to Functional Neuroanatomy*. Academic Press, 2010. ISBN 9780123738899. [101]

J. Weeds and D. Weir. A general framework for distributional similarity. In *Proceedings of EMNLP*, volume 3, 2003. [77]

P. Weiss and H. B. Hiscoe. Experiments on the mechanism of nerve growth. *Journal of Experimental Zoology*, 107(3):315–395, April 1948. ISSN 0022-104X. [105]

M. T. Wiechert, B. Judkewitz, H. Riecke, and R. W. Friedrich. Mechanisms of pattern decorrelation by recurrent neuronal circuits. *Nature neuroscience*, 13(8):1003–10, August 2010. ISSN 1546-1726. [107]

J. Willander and M. Larsson. Smell your way back to childhood: Autobiographical odor memory. *Psychonomic Bulletin & Review*, 13(2):240–244, April 2006. ISSN 1069-9384. [94]

J. Willander and M. Larsson. Olfaction and emotion: the case of autobiographical memory. *Memory & cognition*, 35(7):1659–63, October 2007. ISSN 0090-502X. [93]

D. C. Willhite, K. T. Nguyen, A. V. Masurkar, C. A. Greer, G. M. Shepherd, and W. R. Chen. Viral tracing identifies distributed columnar organization in the olfactory bulb. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (33):12592–7, August 2006. ISSN 0027-8424. [106, 107, 108, 135]

T. J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science (New York, N.Y.)*, 308 (5723):873–6, May 2005. ISSN 1095-9203. [125]

D. J. Willshaw and C. V. D. Malsburg. How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 194(1117):431–45, November 1976. ISSN 0080-4649. [106]

D. R. Wilson and T. R. Martinez. Improved heterogeneous distance functions. *J. Artif. Int. Res.*, 6(January):1—-34, 1997. [32]

D. A. Wilson and R. J. Stevenson. *Learning to smell: olfactory perception from neurobiology to behavior*. Johns Hopkins Univ Pr, Baltimore, Maryland, 2006. ISBN 0801883687. [92, 101, 110, 114, 120, 123, 134, 151, 162, 163]

R. I. R. Wilson and Z. F. Mainen. Early events in olfactory processing. *Annual review of neuroscience*, 29:163–201, January 2006. ISSN 0147-006X. [100]

P. M. Wise, M. J. Olsson, and W. S. Cain. Quantification of odor quality. *Chemical senses*, 25(4):429–43, August 2000. ISSN 0379-864X. [113, 115]

C. C. Woo and M. I. Leon. Increase in a focal population of juxtaglomerular cells in the olfactory bulb associated with early learning. *The Journal of comparative neurology*, 305 (1):49–56, March 1991. ISSN 0021-9967. [158]

W. Woodford. Available aroma of cocaine, 1981. URL `http://www.google.com/patents?hl=en&lr=&vid=USPAT4260517&id=iDw5AAAAEBAJ&oi=fnd&dq=Available+aroma+of+cocaine&printsec=abstract`. [113]

C. Woolsey and E. Walzl. Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. *Bull Johns Hopkins Hosp*, 71:315–344, 1942. [127]

T. A. Woolsey and H. Van der Loos. The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. The description of a cortical field composed of discrete cytoarchitectonic units. *Brain research*, 17(2):205–42, January 1970. ISSN 0006-8993. [128]

T. D. Wyatt. *Pheromones and animal behaviour: communication by smell and taste*. Cambridge University Press, illustrate edition, 2003. ISBN 0521485266. [94]

T. D. Wyatt. Fifty years of pheromones. *Nature*, 457(7227):262–3, January 2009. ISSN 1476-4687. [94]

T. D. Wyatt. Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology*, 196(10):685–700, October 2010. ISSN 1432-1351. [94]

R. Xu and D. Wunsch. Survey of clustering algorithms. *IEEE transactions on neural networks a publication of the IEEE Neural Networks Council*, 16(3):645–678, 2005. [51]

E. Yaksi and R. I. R. Wilson. Electrical coupling between olfactory glomeruli. *Neuron*, 67 (6):1034–47, September 2010. ISSN 1097-4199. [107]

Z. Yan, J. Tan, C. Qin, Y. Lu, C. Ding, and M. Luo. Precise circuitry links bilaterally symmetric olfactory maps. *Neuron*, 58(4):613–24, May 2008. ISSN 1097-4199. [109]

Y. Yeshurun and N. Sobel. An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Annual review of psychology*, 61:219–41, C1–5, January 2010. ISSN 1545-2085. [115]

M. Yokoi. Refinement of Odor Molecule Tuning by Dendrodendritic Synaptic Inhibition in the Olfactory Bulb. *Proceedings of the National Academy of Sciences*, 92(8):3371–3375, April 1995. ISSN 0027-8424. [107]

I. Yoshida and K. Mori. Odorant category profile selectivity of olfactory cortex neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(34): 9105–14, August 2007. ISSN 1529-2401. [111, 130, 151, 159]

S. L. Youngentob, B. A. Johnson, M. I. Leon, P. R. Sheehe, and P. F. Kent. Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. *Behavioral neuroscience*, 120(6):1337–45, December 2006. ISSN 0735-7044. [123, 131]

C. R. Yu, J. Power, G. Barnea, S. O'Donnell, H. E. V. Brown, J. Osborne, R. Axel, and J. A. Gogos. Spontaneous neural activity is required for the establishment and maintenance of the olfactory sensory map. *Neuron*, 42(4):553–66, May 2004. ISSN 0896-6273. [45, 105]

L. Yu and H. Liu. Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution. *Machine Learning*, 20(2):856, 2003. [29]

S. Zabell. The rule of succession. *Erkenntnis*, 31(2):283–321, 1989. [42]

D. H. Zald and J. V. Pardo. Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 94(8):4119–24, April 1997. ISSN 0027-8424. [93]

M. Zarzo. Psychologic Dimensions in the perception of everyday odors: Pleasantness and Edibility. *Journal of Sensory Studies*, 23(3):354–376, June 2008. ISSN 0887-8250. [115, 136, 141, 152, 153, 154]

M. Zarzo. Hedonic Judgments of Chemical Compounds Are Correlated with Molecular Size. *Sensors*, 11(4):3667–3686, March 2011. ISSN 1424-8220. [114]

M. Zarzo and D. T. Stanton. Identification of latent variables in a semantic odor profile database using principal component analysis. *Chemical senses*, 31(8):713–24, October 2006. ISSN 0379-864X. [116]

M. Zarzo and D. T. Stanton. Understanding the underlying dimensions in perfumers' odor perception space as a basis for developing meaningful odor maps. *Attention, perception & psychophysics*, 71(2):225–47, February 2009. ISSN 1943-3921. [115, 116, 141]

A. Zavada, C. L. Buckley, D. Martinez, J.-P. Rospars, and T. Nowotny. Competition-based model of pheromone component ratio detection in the moth. *PloS one*, 6(2):e16308, January 2011. ISSN 1932-6203. [122]

X. Zhang and S. J. Firestein. Genomics of olfactory receptors. *Results and problems in cell differentiation*, 47:25–36, January 2009. ISSN 0080-1844. [93]

X. Zhang, O. De la Cruz, J. M. Pinto, D. Nicolae, S. J. Firestein, and Y. Gilad. Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome biology*, 8(5):R86, January 2007a. ISSN 1465-6914. [102]

X. Zhang, X. Zhang, and S. J. Firestein. Comparative genomics of odorant and pheromone receptor genes in rodents. *Genomics*, 89(4):441–50, April 2007b. ISSN 0888-7543. [99, 102]

J. Zhou and H. Peng. Automatic recognition and annotation of gene expression patterns of fly embryos. *Bioinformatics*, 23(5):589–596, 2007. [25, 43]

D.-J. Zou, P. Feinstein, A. L. Rivers, G. a. Mathews, A. Kim, C. A. Greer, P. Mombaerts, and S. Firestein. Postnatal refinement of peripheral olfactory projections. *Science (New York, N.Y.)*, 304(5679):1976–9, June 2004. ISSN 1095-9203. [105]

Z. Zou, F. Li, and L. B. Buck. Odor maps in the olfactory cortex. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(21):7724–7729, 2005. [111]

Z. Zou, L. F. Horowitz, J.-P. P. Montmayeur, S. Snapper, and L. B. Buck. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature*, 414(7183):173–9, March 2008. ISSN 1476-4687. [111]

H. Zwaardemaker. About odour-affinity, based on experiments of J. Hermanides. *KNAW, Proceedings*, (September):90–97, 1909. [114]

H. Zwaardemaker. The Electrical Phenomenon in Smell-mixtures. *van Weteschappen Proceedings Series B Physical*, (September):551–555, 1917. [112, 113]