Exam in Machine Learning
DD2431

2007-12-20, kl 14.00 – 19.00

Aids allowed: *calculator*, *language dictionary*.

A Questions for pass or fail

Note: To pass the exam you must give the correct answer on *all* questions in this section. Be very careful not to make any unnecessary mistakes here.

A-1 Hypotheses Space

What is the size of the Hypotheses Space (H) for a single layered ANN with three inputs and two outputs?

a) 2

b) 3

c) 8

d) 2^6

e) ∞

A-2 Genetic Algorithms

Which three operations are central for a Genetic Algorithm?

a) Inversion — Reducing the population size with time

b) Selection — Preserving the best individuals

c) Restriction — Preventing inconsistent combinations

d) Crossover — Combining parts of good individuals

e) Mutation — Random changes

f) Postponing — Avoiding changing improving individuals

Note: Answer with *all three* correct items.

A-3 Bayes Classifier

A *naive Bayes classifier* is characterized by the following (choose exactly one answer):

a) All class values are considered to have the same prior probability.

b) All data attributes are considered conditionally independent given the class.

c) One data attribute is selected randomly and used for classification.

d) The classifier maintains a model of the joint likelihood of attributes given the class.

A-4 Reinforcement Learning

In the standard setting for reinforcement learning, what information does the *agent receive* from the environment in each time step?

a) State

b) Q-value

c) Value

d) Reward

e) Policy

f) Discount

Note: Answer with *all* correct alternatives.

A-5 Bias

An inductive bias preferring *simpler hypotheses* tends to

a) Improve generalization for unseen examples

b) Improve accuracy for training examples

c) None of the above

d) Both of the above

A-6 Hypothesis Order

Which hypothesis is most *general*:

a) Rainy

b) Rainy $\wedge$ Windy

B Questions for higher grades

Preliminary number of points required for different grades:

$$\begin{aligned} 0 \leq p < 6 &\rightarrow E \\ 6 \leq p < 12 &\rightarrow D \\ 12 \leq p < 16 &\rightarrow C \\ 16 \leq p < 20 &\rightarrow B \\ 20 \leq p \leq 24 &\rightarrow A \end{aligned}$$

B-1 MAP and ML

(3)

Describe the difference between Maximum A Posteriori (MAP) and Maximum Likelihood (ML) estimate. How are they defined? When are they used? Give an example (no formulas) where the MAP and ML estimates differ.

B-2 AdaBoost

(3)

Describe AdaBoost. What is the requirement on the used weak classifier to guarantee convergence?

B-3 Candidate-Elimination

(3p)

CANDIDATE-ELIMINATION can be used to find the set of consecutive integers, consistant with the training examples. Each training example is a single integer, together with information about if it is in the interval or not.

Show, for each new training example, what the sets S and G will contain during training with these examples:

$$< 1, - >, < 5, - >, < 9, - >, < 6, + >$$

The examples are presented in this order (from left to right).

B-4 Christmas Bells

(3)

For a computer application she is developing, Maria needs the sound of Christmas bells, but she is concerned about potential problems with copyright and prefers to use a synthetic sound rather than recording real bells. She come up with the idea of using a *genetic algorithm* to find a time series (a short wave file) which produces the most beautiful Christmas bell sound. Marias idea is to let people visiting her web portal listen to two sound samples and click on the one that sounds best.

Describe how you would implement Marias idea. In particular, describe:

a) What constitutes individuals and how are they represented in the form of chromosomes?

b) How will you generate new individuals?

c) How will you decide which individuals to preserve?

B-5 Restaurant food

(3)

Arne is somewhat picky with what he eats and often finds the food they serve at the local restaurant un-eatable. He has noted that there are two independent reasons for this: sometimes the food is too cold and sometimes it contains too much salt. Arne estimates the probability that the food is too cold to be 20%, and the probability that it is too salt to 30%.

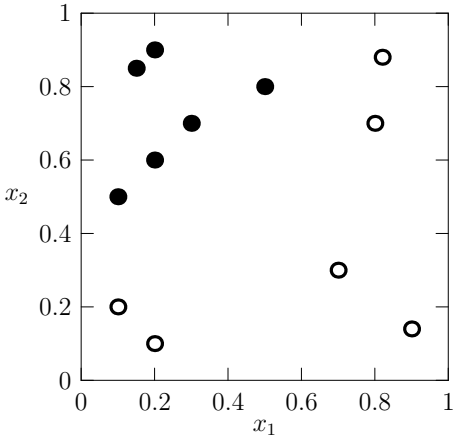
a) How unpredictable is the situation that the food is un-eatable, i.e. too cold or too salt (or both), calculated as an entropy measured in bits?

b) Arne has figured out a (somewhat risky) way of checking the temperature of the food before ordering. What is the expected information gain from finding out if the food is too cold?

B-6 Neural Network

(3)

Manually calculate the weights and threshold value of a single layered neural network so that all the points in the figure are correctly classified. Filled circles should give the output 1, open circles should give 0. Clearly show how the output is computed from the input with a formula where your values are included.



B-7 VC-dimension

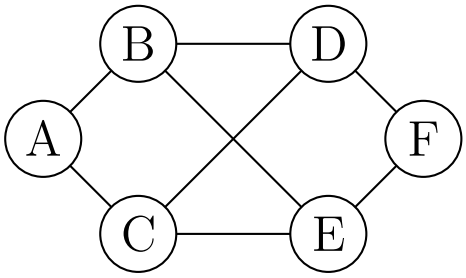
(3)

Consider a *concept learning task* where each example is a point in $\mathcal{R}^2$ (a two dimensional plane). What is the VC-dimension for a hypotheses space where *each hypothesis is a triangle* where points inside the triangle are considered part of the concept? The motivation for you answer is important.

B-8 Labyrinth

(3)

Consider the following "labyrinth" where the labelled nodes denote positions. Position F is the *goal state* where you exit the labyrinth.



Given that each move (along any of the edges in the graph) gives a reward of -1 (i.e. a punishment), what is the value of being in each of the positions $A, B, \ldots, F$ when following an optimal policy? Use a discount factor (γ) of 0.9 and the normal definition of "value" used in reinforcement learning.

As usual, you must show how you arrived at your result.

Good Luck!