Course: DD2427 - Exercise Set 5

In this set of exercises you will examine classifiers based on the nearest neighbour rule.

Exercise 1: *Nearest neighbour decision boundary*

Assume that you have one feature vector $\mathbf{x}_1 = (x_1, y_1)$ from class ω_1 and another feature vector $\mathbf{x}_2 = (x_2, y_2)$ from class ω_2. These two $2d$ points can be used to define a very simple 1NN (nearest neighbour) classifier to distinguish between the classes ω_1 and ω_2. Show that the decision boundary induced by this classifier is a straight line.

(**HINT**: When will a query point $\mathbf{x} = (x, y)$ be classified as class ω_1 ? And similarly when will it be classified as class ω_2 ? And finally, when is the query point not clearly belonging to either class. This latter situation occurs at the decision boundary.)

Exercise 2: *Nearest neighbour example*

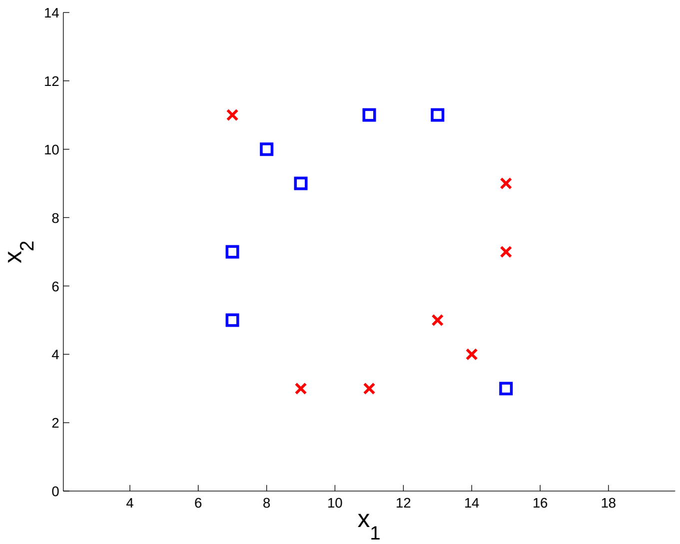

Figure 1: The training data for simple 1-NN classifier.

Next consider we have two attributes $x_1, x_2 \in \mathbb{R}$ and a class label $y \in \{0, 1\}$ and a training set that looks like:

Class 1 points:

$$\{(7,11),(15,9),(15,7),(13,5),(14,4),(9,3),(11,3)\}$$

Class 2 points:

$$\{(11,11),(13,11),(8,10),(9,9),(7,7),(7,5),(15,3)\}$$

Graphically, these points appear as in figure 1. **Draw the decision boundaries for 1-nearest neighbour classification on this data.** Assume the classifier uses a Euclidean distance metric and outputs either `1` or `2`.

The boundary does not need to be exact - a print out of figure 1 with boundaries drawn approximately is sufficient. Make sure to label the regions with the classification label that would be given.

Exercise 3: *Nearest neighbour example II*

One of the problems with k-nearest neighbor classification is selecting a value for k. Say you are given the data set shown in figure 2 as training data from which a kNN classifier is constructed. Assume the classifier uses a Euclidean distance metric and outputs either `1` or `2`.

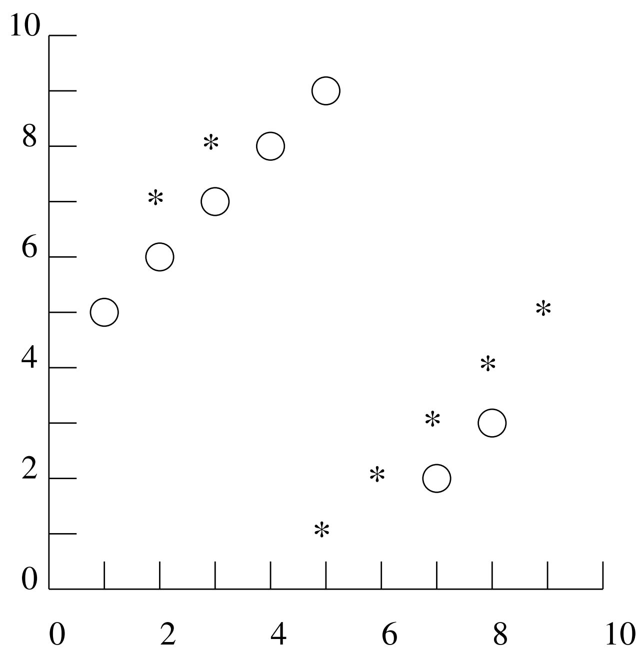

Figure 2: Training data for a kNN classifier.

Questions Why might using too large a value of k be bad for this data set? Why might using too small a value be bad for this data set?

For the lecture: 13th April

Bring the following:

- *Written solutions to Exercises 1 to 3.*

Exercise 4: *Error bound for 1-nearest neighbour classifier (***optional***)*

A classic paper of Cover and Hart [1] from 1967 shows that, as the amount of training data approaches infinity, the error rate of 1-nearest neighbour classifier is at most twice the Bayes-optimal error rate. In this exercise you will go through the proof for the case of binary classification with real-valued inputs.

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be the training examples where each $\mathbf{x}_i \in \mathbb{R}$ and $\{\theta_1, \theta_2, \ldots, \theta_n\}$ be their corresponding class labels, where each $\theta_i \in \{\omega_1, \omega_2\}$. Let $p(\mathbf{x} \,|\, \omega_i)$ be the true class conditional probability distribution for points in class ω_i and $P(\omega_i)$ is the prior probability distribution for each class.

Error of the Bayes' classifier

From Lecture 3 we know that the classifier which minimizes the probability of error is defined and referred to as the Bayes' classifier:

$$\omega_m = \arg\max_i P(\omega_i \,|\, \mathbf{x}) = \arg\max_i p(\mathbf{x} \,|\, \omega_i)\, P(\omega_i)$$

Denote by $P_B(\text{error})$ the error associated with the Bayes' classifier. It can be computed as follows via $P_B(\text{error} \,|\, \mathbf{x})$:

$$P_B(\text{error} \,|\, \mathbf{x}) = 1 - P_B(\text{correct} \,|\, \mathbf{x}) = 1 - P(\omega_m \,|\, \mathbf{x})$$

The error of the Bayes classifier, the best one can do, is then:

$$P_B(\text{error}) = \int_{\mathbf{x}} P_B(\text{error}|\, \mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} = \int_{\mathbf{x}} (1 - P(\omega_m \,|\, \mathbf{x}))\; p(\mathbf{x})\, d\mathbf{x}$$

Error of the 1-NN classifier

Now let's investigate the 1-NN (nearest neighbour) classifier and its probability of error. Suppose that the true class of an unseen point $\mathbf{x}$ is θ and that $\mathbf{x}' \in \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is $\mathbf{x}$'s nearest neighbour from the training data.

For the nearest neighbour classifier the probabiliy of error given $\mathbf{x}$ and its nearest neighbour $\mathbf{x}'$ is:

$$
\begin{aligned}
P_N(\text{error} \,|\, \mathbf{x}, \mathbf{x}') &= 1 - P_N(\text{correct} \,|\, \mathbf{x}, \mathbf{x}') \\
&= 1 - \sum_{i=1}^{2} P(\theta = \omega_i, \theta' = \omega_i \,|\, \mathbf{x}, \mathbf{x}') \\
&= 1 - \sum_{i=1}^{2} P(\theta = \omega_i \,|\, \mathbf{x}, \mathbf{x}')\, P(\theta' = \omega_i \,|\, \mathbf{x}, \mathbf{x}'), \text{ } \mathbf{x} \text{ and } \mathbf{x}' \text{ drawn independently} \\
&= 1 - \sum_{i=1}^{2} P(\theta = \omega_i \,|\, \mathbf{x})\, P(\theta' = \omega_i \,|\, \mathbf{x}')
\end{aligned}
$$

The probability of error for the nearest neighbour classifier given $\mathbf{x}$ is:

$$P_N(\text{error} \,|\, \mathbf{x}) = \int_{\mathbf{x}'} P_N(\text{error} \,|\, \mathbf{x}, \mathbf{x}')\, p(\mathbf{x}' \,|\, \mathbf{x})\, d\mathbf{x}'$$

Error of the 1-NN classifier as training data tends towards ∞

As the number of training examples tends towards infinity that is $n \to \infty$ then $\mathbf{x}' \to \mathbf{x}$ and $p(\mathbf{x}' \,|\, \mathbf{x}) \to \delta(\mathbf{x}' - \mathbf{x})$. Thus as $n \to \infty$

$$P_N(\text{error} \,|\, \mathbf{x}, \mathbf{x}') = 1 - \sum_{i=1}^{2} P(\omega_i \,|\, \mathbf{x})\, P(\omega_i \,|\, \mathbf{x}') \longrightarrow 1 - \sum_{i=1}^{2} P(\omega_i \,|\, \mathbf{x})^2$$

and

$$P_N(\text{error} \,|\, \mathbf{x}) = \int_{\mathbf{x}'} P_N(\text{error} \,|\, \mathbf{x}, \mathbf{x}')\, p(\mathbf{x}' \,|\, \mathbf{x})\, d\mathbf{x}' \quad \longrightarrow 1 - \sum_{i=1}^{2} P(\omega_i \,|\, \mathbf{x})^2 \qquad (1)$$

Let ω_m be the prediction of the class of $\mathbf{x}$ by the Bayes classifier. Then let ω_l be the other class and thus

$$P(\omega_l \,|\, \mathbf{x}) = 1 - P(\omega_m \,|\, \mathbf{x}) = P_B(\text{error} \,|\, \mathbf{x})$$

Therefore

$$P(\omega_m \,|\, \mathbf{x}) = 1 - P_B(\text{error} \,|\, \mathbf{x}) \qquad (2)$$

Use equations (1) and (2) to express $P_N(\text{error} \,|\, \mathbf{x})$ in terms of $P_B(\text{error} \,|\, \mathbf{x})$. Then express $P_N(\text{error})$ in terms of $P_B(\text{error})$ and deduce that $P_N(\text{error}) \leq 2\, P_B(\text{error})$ for the case of infinite training data.

References

[1] T Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, pages 21–27, January 1967.