MACHINE LEARNING

CLASSIFICATION I

ACADEMIC YEAR 2023/2024

QUEEN MARY UNIVERSITY OF LONDON

Exercises

Exercise $\sharp$1. Given a linear classifier $\boldsymbol{w}$ and a sample $\boldsymbol{x}_i$, the quantity $\boldsymbol{w}^T\boldsymbol{x}_i$ can be interpreted as the distance from the sample to the boundary. Furthermore,

- If $\boldsymbol{w}^T\boldsymbol{x}_i > 0$, the sample is in one decision region.
- If $\boldsymbol{w}^T\boldsymbol{x}_i < 0$, the sample is in the other decision region.

Some textbooks use the notion of margin as a distance from a sample to the boundary, such that it is positive if the sample is correctly classified and negative if it is misclassified. Accordingly, they seek to find the classifier that maximises the margin. Suggest a numerical label y_i so that the quantity $m_i = y_i\left[\boldsymbol{w}^T\boldsymbol{x}_i\right]$ defines a margin.

Exercise $\sharp$2. Consider the simple dataset shown in Figure 1, consisting of three samples belonging to class $\bigcirc$ and three samples belonging to class $\bigcirc$ in a 2D predictor space with attributes x_A and x_B.

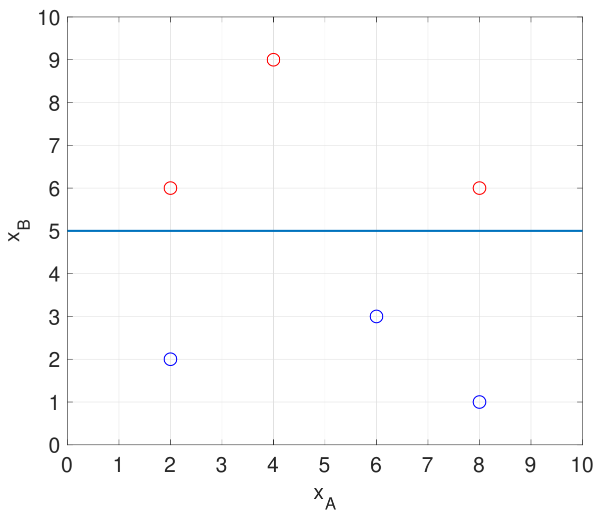

Figure 1: Simple dataset and linear boundary

Assume that we use a classifier whose boundary is the straight line shown in Figure 1.

- Find the coefficients $\boldsymbol{w}$ for the equation $\boldsymbol{w}^T\boldsymbol{x} = 0$ representing the linear boundary of the classifier, where $\boldsymbol{x} = [1, x_A, x_B]$ is the extended predictor vector. Are these coefficients unique?
- Select two samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ on the classifier boundary and show that $\boldsymbol{w}^T\boldsymbol{x}_1 = 0$ and $\boldsymbol{w}^T\boldsymbol{x}_2 = 0$.
- For every sample $\boldsymbol{x}_i$ belonging to class $\bigcirc$, compute the quantity $\boldsymbol{w}^T\boldsymbol{x}_i$ and compare its value with the distance from the sample $\boldsymbol{x}_i$ to the boundary.
- Carry the previous comparison for every sample $\boldsymbol{x}_i$ belonging to class $\bigcirc$.
- Given an arbitrary sample $\boldsymbol{x}$, how would our classifier use the result of the computation $\boldsymbol{w}^T\boldsymbol{x}$ to classify it?
- Define a new classifier using the linear boundary with coefficients $\boldsymbol{w}' = k\boldsymbol{w}$, where k is an arbitrary constant. How would this classifier compare with the one defined by $\boldsymbol{w}$?

EXERCISE ♯3. Figure 2 shows a simple dataset in a 2D predictor space with features x_A and x_B. The dataset consists of three samples belonging to class ○ and three samples belonging to class ○. Using the straight line shown in Figure 2 as the boundary of our linear classifier, repeat the steps described in Exercise 2 for this new scenario.

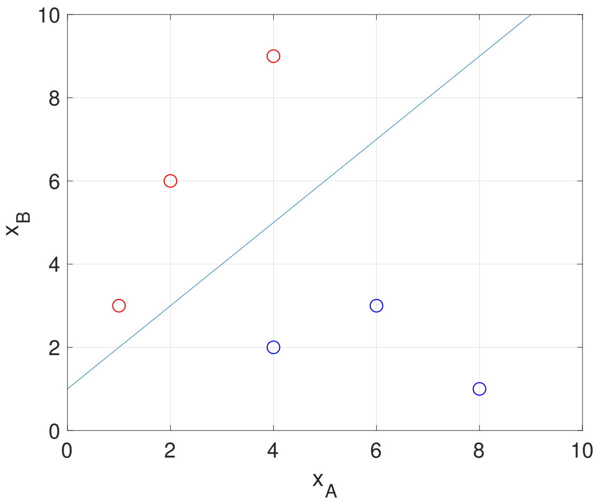

Figure 2: Simple dataset and linear boundary

EXERCISE ♯4. Figure 3 shows four samples belonging to a dataset with predictors x_A, x_B and x_C. As you can see, two samples belong to class • and the other two samples to class •.

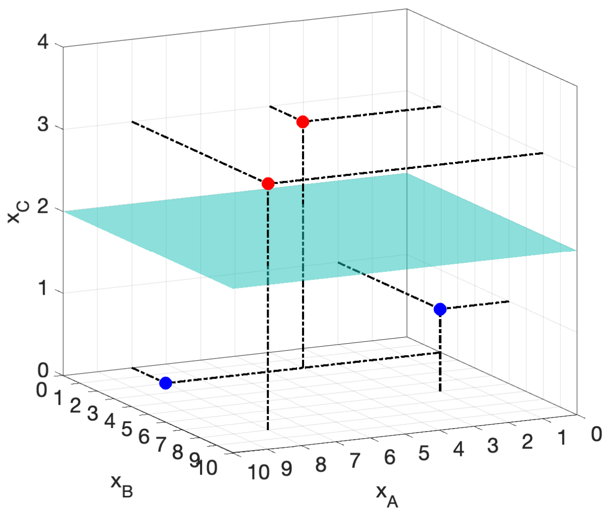

Figure 3: Simple dataset and linear boundary

Consider the linear classifier represented by the surface shown in Figure 3. Repeat all the steps described in Exercise 2 for this new scenario. Note that the extended vector $\boldsymbol{x}$ should now be defined as $\boldsymbol{x} = [1, x_A, x_B, x_C]$, and the coefficients vector describing the classifier's linear boundary will need to be redefined accordingly.

EXERCISE ♯5. Consider a linear classifier defined by the coefficients vector $\boldsymbol{w}$, where samples $\boldsymbol{x}_i$ such that $\boldsymbol{w}^T\boldsymbol{x}_i \geq 0$ are labelled as ◯ (otherwise, they are labelled as ◯). A convenient way to quantify the linear classifier's certainty that a sample $\boldsymbol{x}_i$ belongs to the class ◯ is to use the logistic function and compute the value:

$$p(\boldsymbol{x}_i) = \frac{e^{\boldsymbol{w}^T\boldsymbol{x}_i}}{1 + e^{\boldsymbol{w}^T\boldsymbol{x}_i}}$$

- Show that $p(\boldsymbol{x}_i)$ is 0.5 for samples on the boundary and that as samples move away from the boundary, either $p(\boldsymbol{x}_i) \to 0$ or $p(\boldsymbol{x}_i) \to 1$. How would you interpret these numerical values?
- Obtain the likelihood $L(\boldsymbol{w})$ using the classifier $\boldsymbol{w}$ and the dataset defined in Exercise 2.
- Create a new classifier $\boldsymbol{w}'$ by moving the boundary of the previous classifier one unit of x_B down, i.e. the new classifier is defined by the boundary $x_B = 4$. Obtain the likelihood $L(\boldsymbol{w}')$ of the new classifier using the dataset shown in Figure 1.
- Which classifier is better, the one defined by $\boldsymbol{w}$ or the one defined by $\boldsymbol{w}'$?

EXERCISE ♯6. Figure 4 shows a dataset consisting of samples belonging to classes • and • in a predictor space with attributes x_A and x_B.

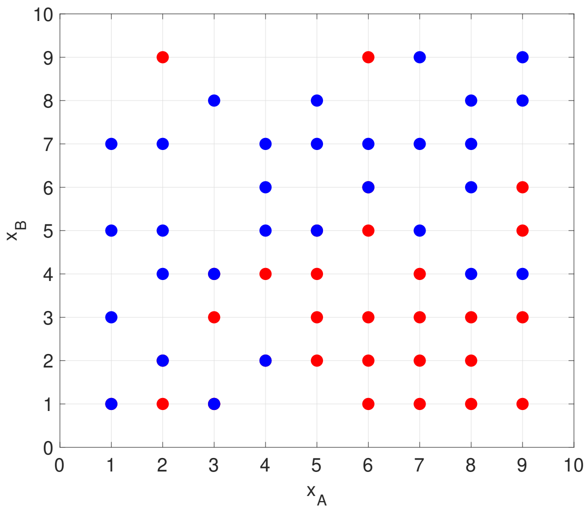

Figure 4

- Sketch the boundaries of three kNN classifiers, where $k = 1, 3$ and 7 respectively. How does the boundary change as k increases? What would the boundary be for $k = 53$?
- Sketch the boundary of a kNN classifier, where $k = 2$ and 4. Why is this choice problematic?