Machine Learning

Classification II

Academic year 2023/2024

Queen Mary University of London

EXERCISES

EXERCISE ♯1. Consider the following dataset, where x will be used as a predictor and y as a label:

x	y
-2	A
-1	A
0	A
1	A
2	A
1	B
2	B
3	B
4	B
5	B

We will assume that the value of the predictor of x is distributed following a Gaussian distribution for both classes A and B. In other words, the class densities are Gaussian.

- Estimate the parameters of the Gaussian class densities.
- Build a Bayes classifier for the previous dataset.
- Build a new Bayes classifier with the same likelihoods but different priors, namely $P_A = 0.1$ and $P_B = 0.9$.

EXERCISE ♯2. Figure 1 shows a dataset consisting of samples belonging to three classes ●, ● and ● in a predictor space with features x_A and x_B.

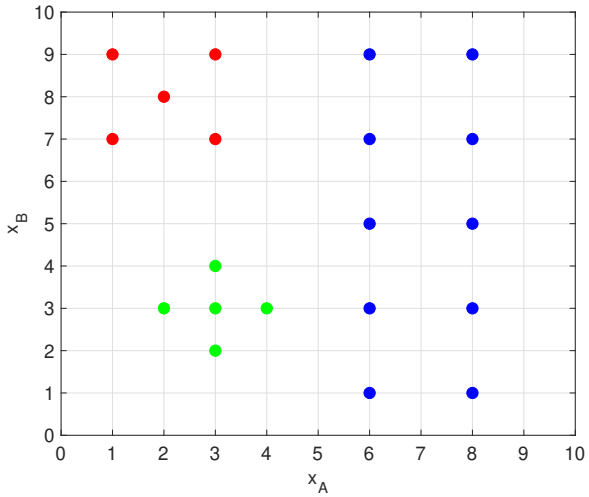

Figure 1

Assuming Gaussian class densities and that x_A and x_B are independent (*naive assumption*), define a Bayes classifier and sketch its boundaries.

EXERCISE ♯3. Figure 2 shows a dataset consisting of samples belonging to classes • and • in a predictor space with features x_A and x_B.

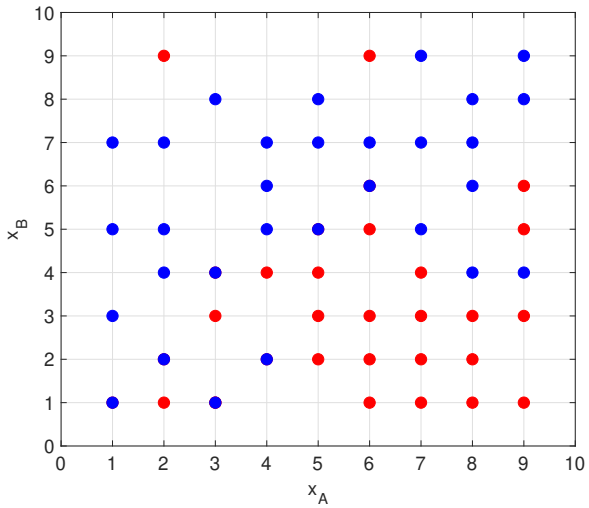

Figure 2

Given the dataset shown in Figure 2, obtain the confusion matrix of the classifiers defined by the boundaries $x_B = 0.5$, $x_B = 1.5$, $x_B = 3.5$, $x_B = 5.5$, $x_B = 7.5$ and $x_B = 9.5$. Assume that samples above each boundary as classified as •, and below as •. Use the resulting rates to sketch the ROC curve of the family of classifiers $x_B = c$, where c is the calibration parameter.

EXERCISE ♯4. Repeat the previous exercise for the family of classifiers defined by $x_B = x_A + c$, where c is the calibration parameter. Obtain the confusion matrix for the boundaries defined by the values $c = -8.5, -4.5, -1.5, 1.5, 4.5, 8.5$ and compare the estimated ROC curve with the ROC curve obtained in the previous exercise. Which family of classifiers represent better the distribution of data?