PRINCIPLES OF MACHINE LEARNING

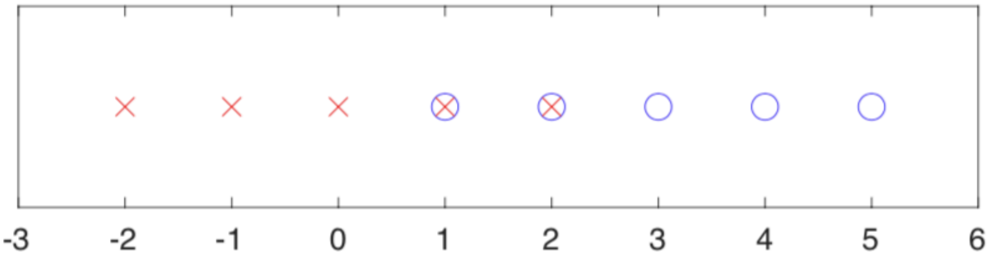
CLASSIFICATION II

ACADEMIC YEAR 2021/2022

QUEEN MARY UNIVERSITY OF LONDON

SOLUTIONS

EXERCISE ♯1 (SOL): Let's start plotting the dataset (we will use symbol X for class A and O for class B):



Both class overlap. In fact two samples have the same predictor and different labels.

- A Gaussian distribution has two parameters, namely the mean μ and standard deviation σ. The estimator for the mean is:

$$\hat{\mu} = \frac{1}{N}\sum_i x_i,$$

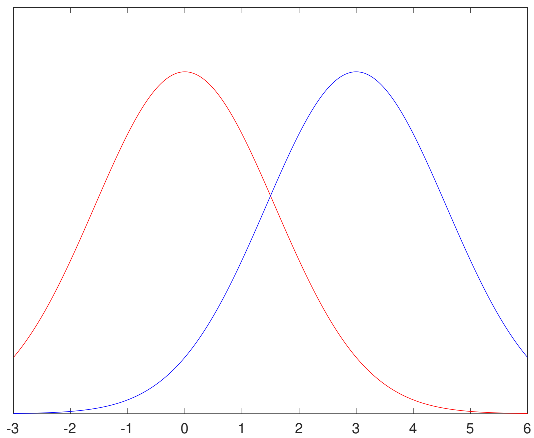
The estimator for the variance can be obtained as the square root of the estimator of the variance σ^2. There are two estimators for the variance, one biased and another unbiased:

$$\hat{\sigma^2} = \frac{1}{N}\sum_i (x_i - \hat{\mu})^2 \quad (\textit{biased}) \quad \text{or} \quad \hat{\sigma^2} = \frac{1}{N-1}\sum_i (x_i - \hat{\mu})^2 \, (\textit{unbiased})$$

Using the estimator of the mean and the square root of the unbiased estimator of the variance we get for each class:

$$\begin{aligned}
\mu_A &= (-2-1+0+1+2)/5 = 0 \\
\sigma_A &= \sqrt{\left((-2-0)^2 + (-1-0)^2 + (0-0)^2 + (1-0)^2 + (2-0)^2\right)/4} = 1.58 \\
\mu_B &= (1+2+3+4+5)/5 = 3 \\
\sigma_B &= \sqrt{\left((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2\right)/4} = 1.58
\end{aligned}$$

Note that both classes have the same standard deviation. If we plot them, we get:



- Given a sample x_i, the Bayes classifier compares the posterior probabilities $P(A|x_i)$ and $P(B|x_i)$ to classify it:

$$\frac{P(A|x_i)}{P(B|x_i)} > 1 \quad \rightarrow \quad \hat{y}_i = A$$
$$\frac{P(A|x_i)}{P(B|x_i)} < 1 \quad \rightarrow \quad \hat{y}_i = B$$

  Using Bayes rule, we can express the posterior probabilities in terms of the priors $P(A)$ and $P(B)$ and the class densities $p(x|A)$ and $p(x|B)$. The class densities are Gaussian and have the same standard deviation (as in linear discriminant analysis) and the priors are $P(A) = 0.5$ and $P(B) = 0.5$. The classifier is then:

$$\frac{P(A|x_i)}{P(B|x_i)} = \frac{P(A)P(x_i|A)}{P(B)P(x_i|B)} = \frac{0.5P(x_i|A)}{0.5P(x_i|B)} = \frac{P(x_i|A)}{P(x_i|B)} > 1 \quad \rightarrow \quad \hat{y}_i = A$$
$$\frac{P(A|x_i)}{P(B|x_i)} = \frac{P(A)P(x_i|A)}{P(B)P(x_i|B)} = \frac{0.5P(x_i|A)}{0.5P(x_i|B)} = \frac{P(x_i|A)}{P(x_i|B)} < 1 \quad \rightarrow \quad \hat{y}_i = B$$

- If the priors are $P(A) = 0.1$ and $P(B) = 0.9$ instead and the class densities (also known as likelihoods) are the same, we get:

$$\frac{0.1P(x_i|A)}{0.9P(x_i|B)} = \frac{P(x_i|A)}{9P(x_i|B)} > 1 \quad \text{or} \quad \frac{P(x_i|A)}{P(x_i|B)} = \frac{P(x_i|A)}{9P(x_i|B)} > 9 \quad \rightarrow \quad \hat{y}_i = A$$
$$\frac{0.1P(x_i|A)}{0.9P(x_i|B)} = \frac{P(x_i|A)}{9P(x_i|B)} < 1 \quad \text{or} \quad \frac{P(x_i|A)}{P(x_i|B)} = \frac{P(x_i|A)}{9P(x_i|B)} < 9 \quad \rightarrow \quad \hat{y}_i = B$$

EXERCISE ♯2 (SOL): A Gaussian distribution in a 2D predictor space has two parameters, namely the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}$$

If predictors x_A and x_B are independent, the covariance matrix is diagonal:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{AA} & 0 \\ 0 & \Sigma_{BB} \end{bmatrix}$$

and its diagonal entries are actually the variances of the marginal class densities:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix}$$

Then, for each class we simply need to estimate the parameters of the marginal class densities. In total, there are 6 class densities (3 classes $\times$ 2 predictors). **Note that in this problem the subindices A and B identify each predictor, whereas in the previous problem they were used to identify each class instead**. The means are

$$\boldsymbol{\mu}_{\color{red}\bullet} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$
$$\boldsymbol{\mu}_{\color{blue}\bullet} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$$
$$\boldsymbol{\mu}_{\color{green}\bullet} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

And the covariance matrices:

$$
\begin{aligned}
\Sigma_{\color{red}\bullet} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\Sigma_{\color{blue}\bullet} &= \begin{bmatrix} 1.1 & 0 \\ 0 & 8.9 \end{bmatrix} \\
\Sigma_{\color{green}\bullet} &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}
\end{aligned}
$$

After obtaining the mean and covariance matrix for each class densities, it is convenient to check that the results make sense. The means should correspond to the centre of each class, the variances should describe the spread of samples around their centres.

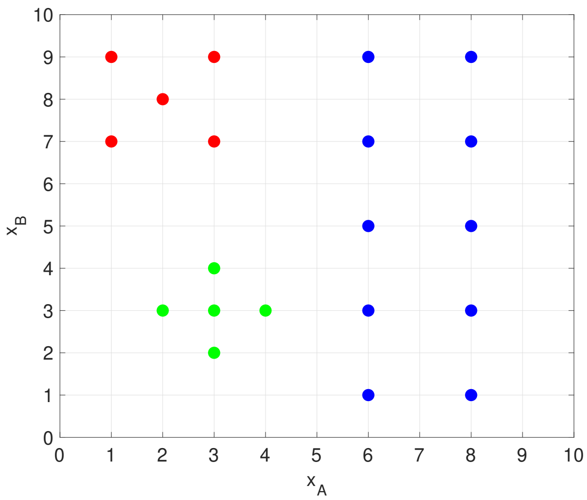


Figure 1

The boundaries of the classifier consist of the points where two or more posterior probabilities are equal. The posterior probabilities can be expressed in terms of the priors and the class densities. Note that in this exercise the priors are $P({\color{red}\bullet})$ = 5/20=1/4, $P({\color{green}\bullet})$ = 5/20=1/4 and $P({\color{blue}\bullet})$ = 10/20=1/2.

Exercise ♯3 (Sol): The dataset consists of 53 samples in a 2D predictor space. There are 30 $\color{blue}\bullet$ samples and 23 $\color{red}\bullet$ samples. In this exercise we are asked to consider several 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 $x_B = 9.5$. We are assuming that samples above each boundary are classified as $\color{blue}\bullet$, and below as $\color{red}\bullet$. A confusion matrix shows the number of correctly and incorrectly classified samples as follows:

		Actual class	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted class	$\color{red}\bullet$	$\color{red}\bullet$ samples labeled $\color{red}\bullet$	$\color{blue}\bullet$ samples labeled $\color{red}\bullet$
	$\color{blue}\bullet$	$\color{red}\bullet$ samples labeled $\color{blue}\bullet$	$\color{blue}\bullet$ samples labeled $\color{blue}\bullet$

We have 6 boundaries, i.e. 6 different classifiers, hence we need to produce a different confusion matrix for each. For each classifier, we need to plot the boundary, count the number of samples correctly and incorrectly classified for each class, and fill in the entries in the confusion matrix.

		Actual	
$x_B = 0.5$		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	0	0
	$\color{blue}\bullet$	23	30

		Actual	
$x_B = 1.5$		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	5	2
	$\color{blue}\bullet$	18	28

		Actual	
$x_B = 3.5$		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	15	5
	$\color{blue}\bullet$	8	25

$x_B = 5.5$		Actual	
		${\color{red}\bullet}$	${\color{blue}\bullet}$
Predicted	${\color{red}\bullet}$	20	14
	${\color{blue}\bullet}$	3	16

$x_B = 7.5$		Actual	
		${\color{red}\bullet}$	${\color{blue}\bullet}$
Predicted	${\color{red}\bullet}$	21	24
	${\color{blue}\bullet}$	2	6

$x_B = 9.5$		Actual	
		${\color{red}\bullet}$	${\color{blue}\bullet}$
Predicted	${\color{red}\bullet}$	23	30
	${\color{blue}\bullet}$	0	0

We will assume ${\color{red}\bullet}$ is the positive class and ${\color{blue}\bullet}$ the negative class. The sensitivity is calculated as:

$$\text{Sensitivity} = \frac{\# {\color{red}\bullet} \text{ samples correctly classified}}{\# {\color{red}\bullet} \text{ samples}}$$

and the specificity as

$$\text{Specificity} = \frac{\# {\color{blue}\bullet} \text{ samples correctly classified}}{\# {\color{blue}\bullet} \text{ samples}}$$

For each classifier, we have the following values of sensitivity and specificity:
$x_B = 0.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{0}{23} = 0 \\ \text{Specificity} &= \frac{30}{30} = 1 \end{aligned}$$

$x_B = 1.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{5}{23} \\ \text{Specificity} &= \frac{28}{30} \end{aligned}$$

$x_B = 3.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{15}{23} \\ \text{Specificity} &= \frac{25}{30} \end{aligned}$$

$x_B = 5.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{20}{23} \\ \text{Specificity} &= \frac{16}{30} \end{aligned}$$

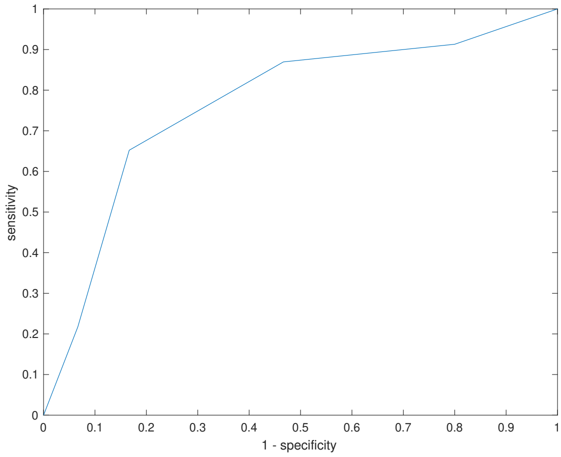
$x_B = 7.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{21}{23} \\ \text{Specificity} &= \frac{6}{30} \end{aligned}$$

$x_B = 9.5$

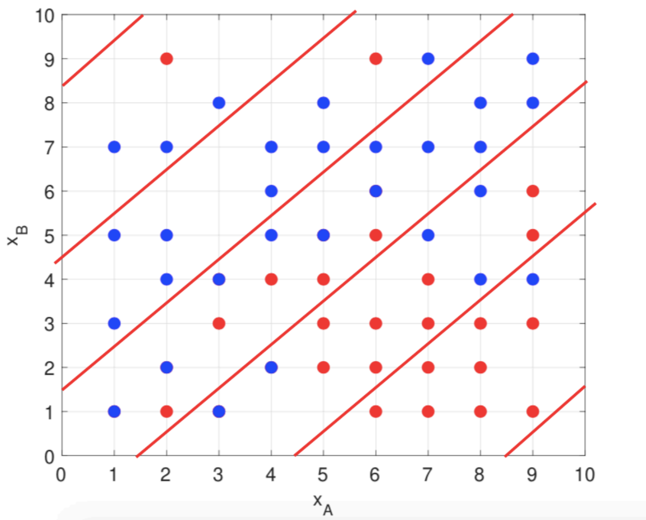
$$\begin{aligned} \text{Sensitivity} &= \frac{23}{23} = 1 \\ \text{Specificity} &= \frac{0}{30} = 0 \end{aligned}$$

Plotting we sensitivity against the 1-specificity we obtain the the ROC curve.



EXERCISE ♯4 (SOL): The boundaries defined by $x_B = x_A + c$ are straight lines with slope 1 and intercept c. Different values of c define a different boundary and a different classifier. Let's assume that samples above each boundary are classified as $\color{blue}\bullet$, and below as $\color{red}\bullet$.

For each classifier (there are 6) we need to plot the boundary, count the number of samples correctly and incorrectly classified for each class, and fill in their confusion matrices.



The confusion matrices are:

$c=-8.5$		Actual	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	0	0
	$\color{blue}\bullet$	23	30

$c=-4.5$		Actual	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	8	1
	$\color{blue}\bullet$	15	29

$c=-1.5$		Actual	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	16	6
	$\color{blue}\bullet$	7	24

$c=1.5$		Actual	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	20	18
	$\color{blue}\bullet$	3	12

$c=4.5$		Actual	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	22	27
	$\color{blue}\bullet$	1	3

$c=8.5$		Actual	
		$\color{red}\bullet$	$\color{blue}\bullet$
Predicted	$\color{red}\bullet$	23	30
	$\color{blue}\bullet$	0	0

The sensitivity and specificity and specificity values are:
$c = -8.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{0}{23} = 0 \\ \text{Specificity} &= \frac{30}{30} = 1 \end{aligned}$$

$c = -4.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{8}{23} \\ \text{Specificity} &= \frac{29}{30} \end{aligned}$$

$c = -1.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{16}{23} \\ \text{Specificity} &= \frac{24}{30} \end{aligned}$$

$c = 1.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{20}{23} \\ \text{Specificity} &= \frac{12}{30} \end{aligned}$$

$c = 4.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{22}{23} \\ \text{Specificity} &= \frac{3}{30} \end{aligned}$$

$c = 8.5$

$$\begin{aligned} \text{Sensitivity} &= \frac{23}{23} = 1 \\ \text{Specificity} &= \frac{0}{30} = 0 \end{aligned}$$

Plotting we sensitivity against the 1-specificity we obtain the the ROC curve. The red curve corresponds to the family $x_B = x_A + c$, the blue to $x_B = c$. The family of classifiers $x_B = x_A + c$ is slightly better than $x_B = c$.