Machine Learning - Homework 2

Pedro Curvo (ist1102716) | Salvador Torpes (ist1102474)

1st Term - 23/24

Dataset

The following dataset will be used for this homework:

D		Input					Output
		y_1	y_2	y_3	y_4	y_5	y_6
Training Observations	x_1	0.24	0.36	1	1	0	A
	x_2	0.16	0.48	1	1	0	A
	x_3	0.32	0.72	0	1	2	A
	x_4	0.54	0.11	0	0	1	B
	x_5	0.66	0.39	0	0	0	B
	x_6	0.76	0.28	1	0	2	B
	x_7	0.41	0.53	0	1	1	B
Testing Observations	x_8	0.38	0.52	0	1	0	A
	x_9	0.42	0.59	0	1	1	B

Table 1: Dataset

1^{st} Question

In order to build the Bayesian classifier for this dataset, we need to compute the class conditional distributions of $\{y_1, y_2\}$, $\{y_3, y_4\}$ and y_5, which are the groups of independent input variables of our dataset as well as the priors.

Priors First of all, we will compute the priors $P(y_6 = A)$ and $P(y_6 = B)$:

$$P(y_6 = A) = \frac{3}{7}$$
$$P(y_6 = B) = \frac{4}{7}$$

Distribution of y_1 and y_2 We are told that $y_1 \times y_2 \in \mathbb{R}$ follows a normal 2D distribution. A multivariate normal distribution of m variables $\vec{x} = \{x_1, x_2, ..., x_m\}$ is defined by its mean vector $\vec{\mu}$ and its covariance matrix Σ:

$$P(\vec{x}|\vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \cdot \Sigma^{-1} \cdot (\vec{x} - \vec{\mu})\right)$$

In our case, we have $m = 2$, $\vec{x} = \{y_1, y_2\}$ and we need to compute two class conditional distributions $p(\vec{x}|y_6 = A)$ and $p(\vec{x}|y_6 = B)$.

Distribution of $\{y_1, y_2\}$ given $y_6 = A$

Considering the training data in table 1 with class $y_6 = A$, we can compute the mean vector $\vec{\mu}$ and the covariance matrix Σ as follows:

$$\vec{\mu} = \begin{bmatrix} \mu_{y_1} \\ \mu_{y_2} \end{bmatrix} = \frac{1}{3} \cdot \begin{bmatrix} 0.24 + 0.16 + 0.32 \\ 0.36 + 0.48 + 0.72 \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.52 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1,y_2} \\ \sigma_{y_1,y_2} & \sigma_{y_2}^2 \end{bmatrix} = \frac{1}{3} \cdot \begin{bmatrix} \sum_{i=1}^{3}(y_{1i} - \mu_{y_1})^2 & \sum_{i=1}^{3}(y_{1i} - \mu_{y_1})(y_{2i} - \mu_{y_2}) \\ \sum_{i=1}^{3}(y_{1i} - \mu_{y_1})(y_{2i} - \mu_{y_2}) & \sum_{i=1}^{3}(y_{2i} - \mu_{y_2})^2 \end{bmatrix} = \begin{bmatrix} 0.0043 & 0.0064 \\ 0.0064 & 0.0224 \end{bmatrix}$$

Now we need to compute both $|\Sigma|$ and Σ^{-1}:

$$|\Sigma| = \det \Sigma = 0.0043 \cdot 0.0224 - 0.0064^2 = 0.00005536$$

$$\Sigma^{-1} =$$

Therefore, we have the normal distribution of $\{y_1, y_2\}$ given $y_6 = A$:

$$P((y_1, y_2)|y_6 = A) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(-\frac{1}{2}((y_1, y_2) - \vec{\mu})^T \cdot \Sigma^{-1} \cdot ((y_1, y_2) - \vec{\mu})\right) =$$

Distribution of $\{y_1, y_2\}$ given $y_6 = B$

Considering the training data in table 1 with class $y_6 = B$, we can compute the mean vector $\vec{\mu}$ and the covariance matrix Σ as follows:

Distribution of y_3 and y_4 The class conditional distributions of y_3 and y_4 come directly from the information in table 1 and they are given by:

$P(y_3 \cap y_4 \mid y_6 = A)$		$y_3 = 0$	$y_3 = 1$
y_4	0	$P(y_3 = 0 \cap y_4 = 0 \mid y_6 = A) = 0$	$P(y_3 = 1 \cap y_4 = 0 \mid y_6 = A) = 0$
	1	$P(y_3 = 0 \cap y_4 = 1 \mid y_6 = A) = \frac{1}{3}$	$P(y_3 = 1 \cap y_4 = 1 \mid y_6 = A) = \frac{2}{3}$

Table 2: Distribution of y_3 and y_4 given $y_6 = A$

$P(y_3 \cap y_4 \mid y_6 = B)$		$y_3 = 0$	$y_3 = 1$
y_4	0	$P(y_3 = 0 \cap y_4 = 0 \mid y_6 = B) = \frac{1}{2}$	$P(y_3 = 1 \cap y_4 = 0 \mid y_6 = B) = \frac{1}{4}$
	1	$P(y_3 = 0 \cap y_4 = 1 \mid y_6 = B) = \frac{1}{4}$	$P(y_3 = 1 \cap y_4 = 1 \mid y_6 = B) = 0$

Table 3: Distribution of y_3 and y_4 given $y_6 = B$

$P(y_5 \mid y_6)$		$y_5 = 0$	$y_5 = 1$	$y_5 = 2$
y_6	A	$P(y_5 = 0 \mid y_6 = A) = \frac{2}{3}$	$P(y_5 = 1 \mid y_6 = A) = 0$	$P(y_5 = 2 \mid y_6 = A) = \frac{1}{3}$
	B	$P(y_5 = 0 \mid y_6 = B) = \frac{1}{4}$	$P(y_5 = 1 \mid y_6 = B) = \frac{1}{2}$	$P(y_5 = 2 \mid y_6 = B) = \frac{1}{4}$

Table 4: Distribution of y_5 given y_6

Distribution of y_5 The class conditional distribution of y_5 is given by:

2nd Question

In order to classify the testing observations, we will need to compute the posterior probabilities.