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

1st Question

In order to build the Bayesian classifier for this dataset, we need to compute the distribution of y_1, y_2, y_3, y_4 and y_5, which are the groups of independent input variables of our dataset.

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 $m = 2$ and we need to compute the mean vector $\vec{\mu}$ and the covariance matrix Σ:

$$\mu = \begin{bmatrix} \mu_{y_1} \\ \mu_{y_2} \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.41 \end{bmatrix}$$

Distribution of y_3 and y_4 The distribution of y_3 and y_4 comes directly from the information of table **??** and is given by:

$P(y_3 \cap y_4)$		y_3	
		0	1
y_4	0	$P(y_3 = 0 \cap y_4 = 0) = \frac{2}{7}$	$P(y_3 = 1 \cap y_4 = 0) = \frac{1}{7}$
	1	$P(y_3 = 0 \cap y_4 = 1) = \frac{2}{7}$	$P(y_3 = 1 \cap y_4 = 1) = \frac{2}{7}$

Distribution of y_5 The distribution of y_5 is given by: