Mahindra University Hyderabad
École Centrale School of Engineering
End Semester-Regular Examination, May 2024
Program: B.Tech Branch: Computation & Mathematics Year: III
Semester: II
Subject: Mathematical Foundations for Machine Learning(MA3219)

Date: 27/05/2024 Start Time: 10.00 AM
Time Duration: 3 Hours Max. Marks: 100

Instructions:

1. Each question carries 20 marks.
2. Justify your answer wherever required. Guesswork will not be considered in evaluation.

Q 1: 20 marks

Use Information gain to construct the decision tree for the following training data.

Weekend	Weather	Parental availability	Wealthy	Decision class
$\mathbf{H_1}$	Sunny	Yes	Rich	Cinema
$\mathbf{H_2}$	Sunny	No	Rich	Tennis
$\mathbf{H_3}$	Windy	Yes	Rich	Cinema
$\mathbf{H_4}$	Rainy	Yes	Poor	Cinema
$\mathbf{H_5}$	Rainy	No	Rich	Home
$\mathbf{H_6}$	Rainy	Yes	Poor	Cinema
$\mathbf{H_7}$	Windy	No	Poor	Cinema
$\mathbf{H_8}$	Windy	No	Rich	Shopping
$\mathbf{H_9}$	Windy	Yes	Rich	Cinema
$\mathbf{H_{10}}$	Sunny	No	Rich	Tennis

Q 2: 20 marks

Consider the sample of six points $X_i \in \mathbb{R}^2$.

$$X = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix} \right\}.$$

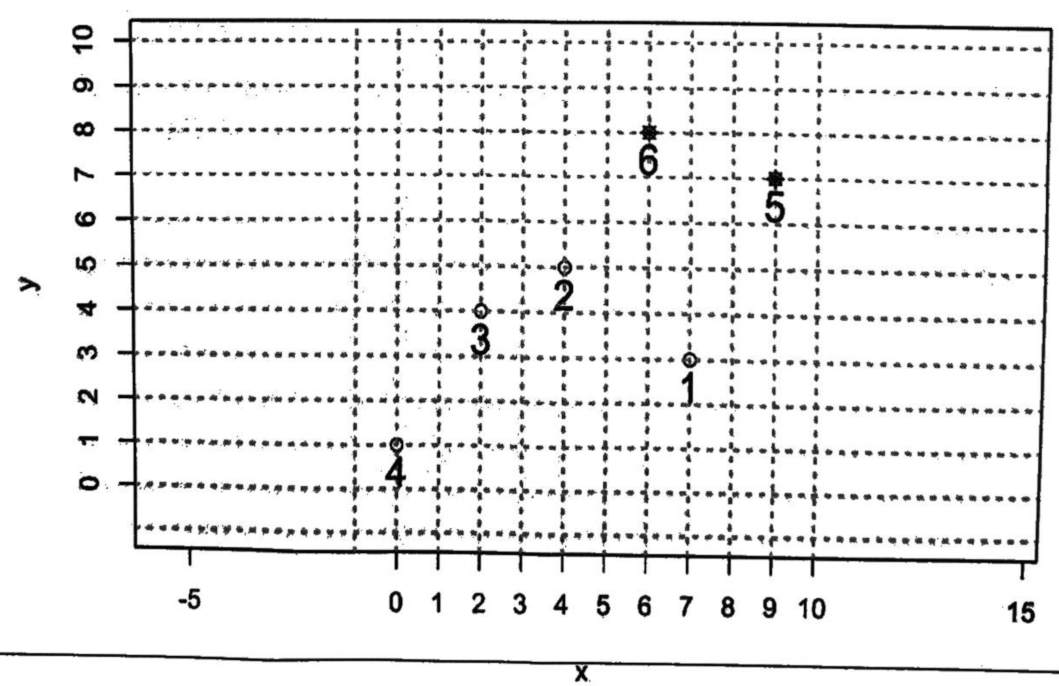
Let's use PCA to reduce its dimension from 2 to 1.

(a) Compute the covariance matrix for the sample points. (Warning: Observe that X is not centered.) Then compute the **unit** eigenvectors, and the corresponding eigenvalues, of the covariance matrix. [12]

(b) Suppose we use PCA to project the sample points onto a one-dimensional space. Which one-dimensional subspace are we projecting onto? For each of the six sample points in X (not the centered version of X!), write the coordinate (in principal coordinate space, not in $\mathbb{R}^2$) to which the point is projected. [8]

Q 3: 20 marks

Perform by-hand k-means algorithm for the points shown in the graph below, with $k = 2$ and with the points $i = 5$ and $i = 6$ as initial centers. Assume that the variables have the same units and no need to scale the data.



Q 4: **20 marks**

(a) Define mean, variance, and standard deviation of a random variable. [8]

(b) Classify the test data {*Red*, *SUV*, *Domestic*} using Naive-Bayes classifier for the dataset shown in table below: [12]

Color	Type	Origin	Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

Q 5: **20 marks**

a) Show that if SVM cost function is written as [12]

$$C(w) = \frac{1}{N}\sum_{i}^{N}\left(\frac{\lambda}{2}\|w\|_2^2 + \max(0, 1 - y_i f(x_i))\right),$$

where $f(x_i) = w^t x_i$, then using steepest descent optimization, w_{t+1} may be learnt from w_t by cycling through the data with the following update rule:

$$\begin{aligned} w_{t+1} &\Longleftarrow (1-\eta\lambda)w_t + \eta y_i x_i \;\; \text{if } y_i w^t x_i < 1, \\ &\Longleftarrow (1-\eta\lambda)w_t \quad \text{otherwise} \end{aligned}$$

where η is learning parameter.

b) Determine the mapping $\phi(x)$ such that the kernel [8]

$$k(x, z) = (c + x^t z)^2 = \phi(x)^t \phi(z),$$

where $x = (x_1, x_2)^t$ and $z = (z_1, z_2)^t$.