U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

June / July 2025 Semester End Main Examinations

Programme: B.E.
Branch: Artificial Intelligence and Machine Learning
Course Code: 23AM5PCIML
Course: Introduction to Machine Learning

Semester: V
Duration: 3 hrs.
Max Marks: 100

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Elaborate the various steps involved in designing the learning system of a checkers problem.	*CO 1*	*PO 1*	**10**
	b)	Apply candidate elimination algorithm and derive the final version space for the training examples: (see table below)	*CO 1*	*PO 2*	**10**

Size	Colour	Shape	Class/Label
Big	Red	Circle	No
Small	Red	Triangle	No
Small	Red	Circle	Yes
Big	Blue	Circle	No
Small	Blue	Circle	Yes

OR

2	a)	Find instance space, Hypothesis Space and Semantically Distinct Hypothesis for the Email classification dataset: (see table below)	*CO 1*	*PO2*	**10**

Email Subject	Email Length	Number of Links	Label
Discount offer	150 words	3	Spam
Meeting invite	300 words	1	Not Spam
Special deal	200 words	2	Spam
Project update	400 words	0	Not Spam
Limited offer	180 words	4	Spam
Team meeting	250 words	1	Not Spam

	b)	Explain the steps of the Find-S algorithm and apply it to solve the given problem. (see table below)	*CO 1*	*PO2*	**10**

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

UNIT - II

3 a) *CO 2* | *PO 2* | **10**

Calculate the entropy, information gain and construct a decision tree for the given dataset using ID3 algorithm.

Instance	a1	a2	a3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

b) *CO 2* | *PO 2* | **10**

i. Write KNN algorithm.
ii. Consider the value of k as 5, apply KNN to the dataset provided and classify the instance {Brightness: 20, Saturation: 35, Class: ?} accordingly.

Brightness	Saturation	Class
40	20	Red
50	50	Blue
60	90	Blue
10	25	Red
70	70	Blue
60	10	Red
25	80	Blue

OR

4 a) *CO 2* | *PO 1* | **10**

Given Positively labeled data points {(3,1), (3, -1), (6,1), (6, -1)} and negatively labelled data points {(1,0), (0,1), (0, -1), (-1,0)}, Classify the same by applying Linear SVM classifier.

b) *CO 2* | *PO 3* | **10**

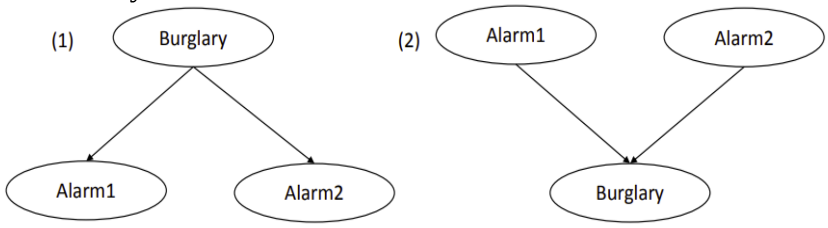
Bob is working on a project to predict housing prices in a city. The dataset Bob has, contains information about various features such as square footage, number of bedrooms, and proximity to amenities. However, Bob notice that the housing market in different neighborhoods behave differently, and a global regression model might not capture these localized variations.

i. Explain how locally weighted regression can be applied to enhance the accuracy of housing price predictions, considering the varying market dynamics in different neighborhoods.
ii. Describe the key steps involved in implementing a locally weighted regression model. Illustrate how the choice of the weighted function and bandwidth parameter can impact the model's performance.
iii. Provide insights on how this localized approach addresses the challenges posed by heterogeneous market conditions across neighborhoods.

UNIT - III

5 a) Derive the equations of the Maximum Likelihood and Least Squared Error Hypothesis. — *CO 2* | *PO 1* | **10**

b) For the Bayesian Network — *CO 2* | *PO 2* | **10**



Assume that:

P(Alarm1) = 0.1
P(Alarm2) = 0.2
P(Burglary | Alarm1, Alarm2) = 0.8
P(Burglary | Alarm1, ¬ Alarm2) = 0.7
P(Burglary | ¬ Alarm1, Alarm2) = 0.6
P(Burglary | ¬ Alarm1, ¬ Alarm2) = 0.5

Find

i. P(Alarm2 | Burglary, Alarm1)
ii. P(Alarm1, Alarm2, Burglary)
iii. P(Alarm1, ¬ Alarm2, Burglary)
iv. P(Burglary, Alarm1)

OR

6 a) The frequency of the features {Yellow, Sweet} required to classify the fruits is summarized: — *CO 2* | *PO 2* | **10**

Fruit	Yellow	Sweet	Total
Mango	350	450	800
Banana	400	300	700
Total	750	750	1500

Using the concept of Naïve Bayes, predict the type of fruit possessing the properties {Yellow, Sweet}

b) With an example, explain EM algorithm in detail. — *CO 2* | *PO 1* | **10**

UNIT - IV

7 a) — *CO 3* | *PO 1* | **10**

i. Illustrate any 2 methods that could be employed to determine the optimal value of k in k-means clustering along with the pros and cons of each method.
ii. Apply Silhouette method for the given data & obtain the optimal value of k.

Point	Cluster Label
P1	1
P2	1
P3	2
P4	2

Point	P1	P2	P3	P4
P1	0	0.10	0.65	0.55
P2	0.10	0	0.70	0.60
P3	0.65	0.70	0	0.30
P4	0.55	0.60	0.30	0

b) Given: — *CO 3* — *PO 1* — **10**

Image ID	Association Tags
1	Beach, Sunshine, Holiday
2	Sand, Beach
3	Sunshine, Beach, Ocean
4	Ocean, People, Beach, Sunshine
5	Holiday, Sunshine

Assume the min_support as 40% and min_confidence as 70%. Apply Apriori algorithm to

i. Find frequent itemset.
ii. Identify strong association rule.

OR

8 a) For the data points, apply single-link hierarchical clustering technique to identify the clusters. Conclude your answer with a dendrogram. — *CO 3* — *PO 2* — **10**

Point	A	B
P1	0.07	0.83
P2	0.85	0.14
P3	0.66	0.89
P4	0.49	0.64
P5	0.80	0.46

b) Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: — *CO 3* — *PO 2* — **10**

A1=(2,10),A2=(2,5),A3=(8,4),A4=(5,8),A5=(7,5),A6=(6,4),A7=(1,2),A8=(4,9).

Suppose that the initial seeds (centers of each cluster) are A1, A4 and A7. Suppose we assign A1, A4 and A7 as the center of each cluster, respectively.

UNIT - V

9 a) A software developer working on the dataset comprising of images of handwritten digits suspect that the traditional feature representations might not capture the underlying manifold structure effectively. — *CO 3* — *PO 3* — **08**

i. How can Locally Linear Embedding (LLE) be employed to enhance the performance of a digit recognition system, particularly in capturing the local relationships and preserving the intrinsic structure within the dataset?
ii. Illustrate the key steps involved in implementing LLE, including the construction of the neighborhood graph and the computation of the low-dimensional embeddings.
iii. What could be the impact of varying the neighborhood size on the quality of the embeddings.
iv. How does LLE address the challenges of nonlinear manifold representations in handwritten digit recognition?

<table>
<tr><td></td><td></td><td>b)</td><td>Determine the 1st principal components of the 2D data points (2,1), (3,5), (4,3), (5,6), (6,7), (7,8) using PCA.</td><td>CO 3</td><td>PO 2</td><td>12</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>10</td><td>a)</td><td>The retail company is keen on enhancing its customer segmentation to tailor marketing campaigns more effectively. Considering the dataset includes demographic information such as age, income, and purchasing behavior metrics.
i. Propose how Linear Discriminant Analysis (LDA) could be applied to identify distinctive customer groups.
ii. Outline the steps you would take to implement LDA in this scenario, emphasizing how the results could guide the marketing team in creating targeted strategies for different customer segments.
Additionally, discuss the potential challenges and limitations associated with using LDA in this context and suggest strategies to mitigate them.</td><td>CO 3</td><td>PO3</td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>Given the data in table, reduce the dimension from 2 to 1 using the Principal Component Analysis algorithm.
<table><tr><th>F</th><th>Ex 1</th><th>Ex 2</th><th>Ex 3</th><th>Ex 4</th></tr><tr><td>X_1</td><td>4</td><td>8</td><td>13</td><td>7</td></tr><tr><td>X_2</td><td>11</td><td>4</td><td>5</td><td>14</td></tr></table></td><td>CO 3</td><td>PO 2</td><td>10</td></tr>
</table>

******