U.S.N.										

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

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

Programme: B.E. **Semester: IV**

Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**

Course Code: 24AM4PCIML **Max Marks: 100**

Course: Introduction to Machine Learning

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

		UNIT - I	CO	PO	Marks
1	a)	Describe any 2 main challenges of Machine Learning.	CO2	PO2	**06**
	b)	Illustrate the version space representation Theorem.	CO2	PO2	**06**
	c)	Using Candidate Elimination algorithm, derive the most specific and generic hypotheses for given data:	CO1	PO1	**08**

Citations	Size	InLibrary	Price	Editions	Buy
Some	Small	No	Affordable	One	No
Many	Big	No	Expensive	Many	Yes
Many	Medium	No	Expensive	Few	Yes
Many	Small	No	Affordable	Many	Yes

		OR			
2	a)	Illustrate List-Then-Eliminate algorithm along with its working procedure.	CO2	PO2	**06**
	b)	Will the Candidate-elimination algorithm converge to correct hypothesis? Justify.	CO2	PO1	**06**
	c)	Apply Find-S algorithm to the given data and derive the corresponding Hypothesis. Explain the process clearly.	CO1	PO1	**08**

hair	body	likesSimon	pose	smile	smart	c(d)
blond	thin	yes	arrogant	toothy	no	1
brown	thin	no	natural	pleasant	yes	0
blond	plump	yes	goofy	pleasant	no	1
black	thin	no	arrogant	none	no	0
blond	plump	no	natural	toothy	yes	0

			UNIT - II			
	3	a)	Apply K-Nearest Neighbor algorithm on the given data with K=3 and determine the target value of the new instance (5, 5). <table><tr><td>**Durability**</td><td>7</td><td>6</td><td>7</td><td>6</td><td>3</td><td>1</td><td>4</td><td>3</td></tr><tr><td>**Strength**</td><td>7</td><td>4</td><td>4</td><td>5</td><td>4</td><td>4</td><td>3</td><td>5</td></tr><tr><td>**Quality**</td><td>Good</td><td>Bad</td><td>Good</td><td>Good</td><td>Bad</td><td>Bad</td><td>Bad</td><td>Bad</td></tr></table>	*CO3*	*PO3*	**10**
		b)	Construct decision tree using the following rules: • (Refund = yes) ➔ No • (Refund = No, Marital Status = {Single, Divorced}, Taxable Income <80K) ➔ No • (Refund = No, Marital Status = {Single, Divorced}, Taxable Income >80K) ➔ yes • (Refund = No, Marital Status = {Married}) ➔ No.	*CO3*	*PO3*	**05**
		c)	Elaborate on the Inductive bias in ID3 search strategy.	*CO2*	*PO1*	**05**
			OR			
	4	a)	Apply the principles of Support Vector Machine (SVM), and derive the best hyperplane possible to classify the given datapoints: (x, y, ClassLabel): (4, 1, +), (4, -1, +), (6, 0, +), (1, 0, -), (0, 1, -), (0, -1, -).	*CO3*	*PO3*	**10**
		b)	Explain how the decision tree learning algorithm performs a search through the hypothesis space. What type of search strategy does it employ, and what are the implications of this choice for completeness and optimality?	*CO2*	*PO1*	**05**
		c)	Elaborate on the mathematical relation that exist between error function and Gradient update in a Locally weighted regression.	*CO1*	*PO1*	**05**
			UNIT - III			
	5	a)	Differentiate between Hard and Soft voting classifiers with an example.	*CO2*	*PO1*	**06**
		b)	Prove that $h_{MDL} = h_{MAP}$ using the concept of Minimum Description Length Principle.	*CO2*	*PO2*	**06**
		c)	Consider a Bayesian Belief Network (BBN) used in a spam email detection system. The network consists of three binary variables: • S: Email is spam • W: Contains the word "WIN" • L: Has a suspicious link The known probabilities are: P(S = true) = 0.2 P(W = true | S=true) = 0.8 P(W= true | S = false) = 0.1 P(L = true | S = true) = 0.9 P(L = true | S = false) = 0.2	*CO1*	*PO2*	**08**

<table>
<tr><td></td><td></td><td></td><td>Answer the following:
i. Given that the word “WIN” is present, what is the probability that the email is spam?
ii. Given that the email contains a suspicious link, what is the probability that it is spam?
iii. Given that both “WIN” is present and a suspicious link is found, what is the probability that the email is spam?</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>6</td><td>a)</td><td>Differentiate between Boosting and Stacking.</td><td>CO2</td><td>PO1</td><td>06</td></tr>
<tr><td></td><td></td><td>b)</td><td>Illustrate the Expectation Maximization (EM) algorithm with all mathematical expressions.</td><td>CO1</td><td>PO1</td><td>06</td></tr>
<tr><td></td><td></td><td>c)</td><td>The total probability of two plant species A and B are P(A)=0.6 and P(B)=0.4 respectively. The conditional probabilities of the features are:
<table><tr><th>Feature</th><th>Value</th><th>P(Value|A)</th><th>P(Value|B)</th></tr><tr><td>Color</td><td>Yellow</td><td>0.7</td><td>0.2</td></tr><tr><td>Shape</td><td>Round</td><td>0.6</td><td>0.3</td></tr><tr><td>Size</td><td>Small</td><td>0.8</td><td>0.4</td></tr><tr><td>Smell</td><td>Yes</td><td>0.3</td><td>0.9</td></tr></table>Use the Naïve Bayes classification technique to predict the species (A or B) for the entity:
{Color=Yellow, Shape=Round, Size=Small, Smell=Yes}</td><td>CO3</td><td>PO3</td><td>08</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - IV</td><td></td><td></td><td></td></tr>
<tr><td></td><td>7</td><td>a)</td><td>Using K-means algorithm, classify the given data points {2, 4, 10, 12, 3, 20, 30, 11, 25} to 2 clusters. Show the intermediate steps until convergence. Start the iteration by considering the data points 4 and 11 as initial cluster centroids.</td><td>CO3</td><td>PO3</td><td>12</td></tr>
<tr><td></td><td></td><td>b)</td><td>Explain any 2 anomaly detection techniques in detail.</td><td>CO2</td><td>PO1</td><td>08</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>8</td><td>a)</td><td>Elaborate on the importance of Support and Confidence in Association Rule Mining.</td><td>CO2</td><td>PO1</td><td>05</td></tr>
<tr><td></td><td></td><td>b)</td><td>Explain the procedural steps of ELBOW method to determine the optimal value of k in the K-means algorithm.</td><td>CO1</td><td>PO1</td><td>05</td></tr>
<tr><td></td><td></td><td>c)</td><td>Classify the given points into various clusters using Hierarchical Agglomerative Single Linkage Clustering technique. Conclude the answer in the form of a dendrogram.
<table><tr><th>Points</th><th>P1</th><th>P2</th><th>P3</th><th>P4</th><th>P5</th><th>P6</th></tr><tr><td>x</td><td>0.4</td><td>0.21</td><td>0.35</td><td>0.26</td><td>0.08</td><td>0.45</td></tr><tr><td>y</td><td>0.53</td><td>0.38</td><td>0.32</td><td>0.19</td><td>0.41</td><td>0.3</td></tr></table></td><td>CO3</td><td>PO3</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - V</td><td></td><td></td><td></td></tr>
<tr><td></td><td>9</td><td>a)</td><td>“Suppose Feature A1 and A2 both achieve 60% accuracy individually. Feature A3 gives only 50% accuracy (same as random guessing in binary classification). However, when A3 is combined with A1, the subset {A1, A3} achieves 90% accuracy.”.</td><td>CO3</td><td>PO3</td><td>08</td></tr>
</table>

<table>
<tr><td></td><td></td><td></td><td>Which category of feature subset selection does the given scenario best suits and why? Support your answer with the appropriate algorithmic procedure and explanation.</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>b)</td><td>Differentiate between feature extraction and feature selection methods of Dimensionality reduction technique.</td><td>CO2</td><td>PO1</td><td>06</td></tr>
<tr><td></td><td></td><td>c)</td><td>Explain the procedural steps of Linear Discriminant Analysis.</td><td>CO2</td><td>PO1</td><td>06</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>Reduce the dimensionality of the given dataset using Principal Component Analysis algorithm.
<table>
<tr><th>Data Point / Feature</th><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td>X1</td><td>4</td><td>8</td><td>13</td><td>7</td></tr>
<tr><td>X2</td><td>11</td><td>4</td><td>5</td><td>14</td></tr>
</table></td><td>CO3</td><td>PO3</td><td>12</td></tr>
<tr><td></td><td></td><td>b)</td><td>How does Locally Linear Embedding (LLE), linearly model local relationships and reduce dimensionality while preserving relationships? Explain with relevant mathematical expressions.</td><td>CO1</td><td>PO1</td><td>08</td></tr>
</table>

******