

--	--	--	--	--	--	--	--	--	--

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

Autonomous Institute Affiliated to VTU

## January / February 2025 Semester End Main Examinations

**Programme: B.E.**

**Semester: VI**

**Branch: Institutional Elective**

**Duration: 3 hrs.**

**Course Code: 24AM6OEIML**

**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.

<b>UNIT - I</b>			<i>CO</i>	<i>PO</i>	<b>Marks</b>
1	a)	Define Machine Learning and illustrate its distinction from traditional programming using block diagrams.	<i>CO1</i>	<i>PO1</i>	<b>5</b>
	b)	Outline the pertinent challenges involved in Machine Learning.	<i>CO1</i>	<i>PO2</i>	<b>5</b>
	c)	Describe the factors influencing step selection in designing a learning system for checkers game, and sketch the final system design.	<i>CO2</i>	<i>PO2</i>	<b>10</b>
<b>OR</b>					
2	a)	A company database named "Hope for Paws" has 10,000 low-resolution images of different dogs taken from a camera and 200,000 images of different dogs gathered by crawling the web. <ul style="list-style-type: none"> <li>i. Elucidate the potential problems that the company may encounter if it decides to build a dog classifier for door camera devices using the above database.</li> <li>ii. How can this problem be rectified?</li> </ul>	<i>CO1</i>	<i>PO1</i>	<b>10</b>
	b)	A data scientist is working with a large dataset containing customer transactions collected over years for a banking sector. The dataset may contain errors, missing values, and inconsistencies. <ul style="list-style-type: none"> <li>i. Explain the importance of data cleaning in constructing predictive models.</li> <li>ii. Describe potential challenges encountered during the data cleaning and ways to overcome them.</li> </ul>	<i>CO1</i>	<i>PO2</i>	<b>10</b>
<b>UNIT - II</b>					
3	a)	Illustrate how the concept learning can be used as the task of searching through a large space of hypothesis.	<i>CO1</i>	<i>PO2</i>	<b>5</b>

**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.

	b)	Apply Find-S algorithm to determine the maximally specific hypothesis for detecting malignant tumors from MRI scans with the provided readings.	CO2	PO3	5																														
		<table border="1"> <thead> <tr> <th>Size</th><th>Shape</th><th>Density</th><th>Tumor</th></tr> </thead> <tbody> <tr> <td>Small</td><td>Regular</td><td>Thin</td><td>Not-Malignant</td></tr> <tr> <td>Medium</td><td>Irregular</td><td>Thick</td><td>Malignant</td></tr> <tr> <td>Large</td><td>Irregular</td><td>Thick</td><td>Malignant</td></tr> </tbody> </table>	Size	Shape	Density	Tumor	Small	Regular	Thin	Not-Malignant	Medium	Irregular	Thick	Malignant	Large	Irregular	Thick	Malignant																	
Size	Shape	Density	Tumor																																
Small	Regular	Thin	Not-Malignant																																
Medium	Irregular	Thick	Malignant																																
Large	Irregular	Thick	Malignant																																
	c)	Apply Candidate Elimination Algorithm for the given loan approval data to derive the version space:	CO2	PO3	10																														
		<table border="1"> <thead> <tr> <th>Ache</th><th>Breathing-Issue</th><th>Cough</th><th>Fatigue</th><th>Fever</th><th>Covid</th></tr> </thead> <tbody> <tr> <td>Severe</td><td>Exists</td><td>Intense</td><td>Present</td><td>High</td><td>Yes</td></tr> <tr> <td>Severe</td><td>Exists</td><td>Intense</td><td>Absent</td><td>High</td><td>Yes</td></tr> <tr> <td>Mild</td><td>Exists</td><td>Light</td><td>Present</td><td>Low</td><td>No</td></tr> <tr> <td>Mild</td><td>Exists</td><td>Intense</td><td>Absent</td><td>High</td><td>Yes</td></tr> </tbody> </table>	Ache	Breathing-Issue	Cough	Fatigue	Fever	Covid	Severe	Exists	Intense	Present	High	Yes	Severe	Exists	Intense	Absent	High	Yes	Mild	Exists	Light	Present	Low	No	Mild	Exists	Intense	Absent	High	Yes			
Ache	Breathing-Issue	Cough	Fatigue	Fever	Covid																														
Severe	Exists	Intense	Present	High	Yes																														
Severe	Exists	Intense	Absent	High	Yes																														
Mild	Exists	Light	Present	Low	No																														
Mild	Exists	Intense	Absent	High	Yes																														
		<b>OR</b>																																	
4	a)	Analyze the effectiveness of the List-then-Eliminate algorithm approach in solving the problem, and enumerate its limitations.	CO1	PO2	6																														
	b)	Write a short note on Version space with respect to Hypothesis space and training data.	CO1	PO2	6																														
	c)	Compare and Contrast Find-S algorithm with Candidate Elimination algorithm.	CO2	PO1	8																														
		<b>UNIT - III</b>																																	
5	a)	Distinguish Linear and Logistic regression techniques.	CO1	PO2	5																														
	b)	Illustrate the working of Linear Support Vector Machine.	CO2	PO2	5																														
	c)	For the given data, Apply KNN with k=5 and classify the new instance (80,40) to a specific class.	CO2	PO3	10																														
		<table border="1"> <thead> <tr> <th>Brightness</th><th>40</th><th>50</th><th>60</th><th>10</th><th>70</th><th>60</th><th>25</th></tr> </thead> <tbody> <tr> <th>Saturation</th><td>20</td><td>50</td><td>90</td><td>25</td><td>70</td><td>10</td><td>80</td></tr> <tr> <th>Class</th><td>Red</td><td>Blue</td><td>Blue</td><td>Red</td><td>Blue</td><td>Red</td><td>Blue</td></tr> </tbody> </table>	Brightness	40	50	60	10	70	60	25	Saturation	20	50	90	25	70	10	80	Class	Red	Blue	Blue	Red	Blue	Red	Blue									
Brightness	40	50	60	10	70	60	25																												
Saturation	20	50	90	25	70	10	80																												
Class	Red	Blue	Blue	Red	Blue	Red	Blue																												
		<b>OR</b>																																	
6	a)	Justify the importance of choosing the right value for k in k-nearest neighbor with an example.	CO1	PO2	4																														
	b)	Researchers in a lab are studying a chemical reaction for a new compound. The data records the mass of the compound over time and it is as follows.	CO2	PO3	8																														
		<table border="1"> <thead> <tr> <th>Time Unit (x)</th><th>5</th><th>7</th><th>12</th><th>16</th><th>20</th></tr> </thead> <tbody> <tr> <th>Mass(y)</th><td>40</td><td>120</td><td>180</td><td>210</td><td>240</td></tr> </tbody> </table>	Time Unit (x)	5	7	12	16	20	Mass(y)	40	120	180	210	240																					
Time Unit (x)	5	7	12	16	20																														
Mass(y)	40	120	180	210	240																														

		Apply the Simple Linear Regression and predict the mass of compound at time unit 10 and 15 respectively.																															
	c)	The dataset of promotion decisions for five employees is given below.  <table border="1"> <thead> <tr> <th>Years of experience</th><th>1</th><th>2</th><th>3</th><th>5</th><th>6</th><th>7</th></tr> </thead> <tbody> <tr> <td>Yes (1) / No (0)</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> </tbody> </table> Apply Logistic Regression with optimizer $z = -10 + 3 * \text{years}$ to: <ol style="list-style-type: none"> <li>Compute the probability of promotion for an employee with 4 years of experience.</li> <li>Determine the minimum years of experience needed for an employee to have over 97% probability of promotion.</li> </ol>	Years of experience	1	2	3	5	6	7	Yes (1) / No (0)	0	0	0	1	1	1	CO2	PO3	8														
Years of experience	1	2	3	5	6	7																											
Yes (1) / No (0)	0	0	0	1	1	1																											
		<b>UNIT - IV</b>																															
7	a)	Illustrate the general structure of decision tree with suitable example.	CO1	PO1	5																												
	b)	Find the entropy for the given probabilities.  <table border="1"> <thead> <tr> <th>P1</th><th>P2</th><th>P3</th><th>P4</th></tr> </thead> <tbody> <tr> <td>0.1</td><td>0.2</td><td>0.3</td><td>0.4</td></tr> </tbody> </table>	P1	P2	P3	P4	0.1	0.2	0.3	0.4	CO2	PO2	5																				
P1	P2	P3	P4																														
0.1	0.2	0.3	0.4																														
	c)	Design a decision tree for the given dataset using the Iterative Dichotomiser (ID3) algorithm.  <table border="1"> <thead> <tr> <th>Instance</th><th>A1</th><th>A2</th><th>Classification</th></tr> </thead> <tbody> <tr> <td>1</td><td>False</td><td>True</td><td>\$</td></tr> <tr> <td>2</td><td>False</td><td>True</td><td>\$</td></tr> <tr> <td>3</td><td>False</td><td>False</td><td>+</td></tr> <tr> <td>4</td><td>True</td><td>False</td><td>\$</td></tr> <tr> <td>5</td><td>True</td><td>True</td><td>+</td></tr> <tr> <td>6</td><td>True</td><td>True</td><td>+</td></tr> </tbody> </table>	Instance	A1	A2	Classification	1	False	True	\$	2	False	True	\$	3	False	False	+	4	True	False	\$	5	True	True	+	6	True	True	+	CO2	PO3	10
Instance	A1	A2	Classification																														
1	False	True	\$																														
2	False	True	\$																														
3	False	False	+																														
4	True	False	\$																														
5	True	True	+																														
6	True	True	+																														
		<b>OR</b>																															
8	a)	With a suitable example, explain the advantages and disadvantages of decision tree learning.	CO1	PO1	10																												
	b)	Highlight the importance of Information Gain in the construction of decision tree.	CO2	PO1	10																												
		<b>UNIT - V</b>																															
9	a)	Illustrate the interactions of Reinforcement Learning components during the learning process.	CO1	PO1	4																												
	b)	Cluster the values (1,3,9,11,2,19,29,10,24) into two groups using k-means clustering with initial centroids $M1 = 3$ and $M2 = 10$ .	CO2	PO3	8																												

	c)	Create a dendrogram by merging clusters based on minimum distance for the data points (17, 21, 24, 41, 26, 42) and update the proximity matrix using hierarchical clustering.	CO2	PO3	<b>8</b>																
		<b>OR</b>																			
10	a)	Compare Supervised and Unsupervised Learning techniques.	CO1	PO2	<b>4</b>																
	b)	Generate clusters for the numbers provided using the Density Based Scan (DBSCAN) algorithm, ensuring each cluster contains at least 4 elements and setting the epsilon ( $\epsilon$ ) parameter to 2.5.  <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td>H</td></tr> <tr> <td>(3,7)</td><td>(4,6)</td><td>(5,5)</td><td>(6,4)</td><td>(7,3)</td><td>(6,2)</td><td>(7,2)</td><td>(8,4)</td></tr> </table>	A	B	C	D	E	F	G	H	(3,7)	(4,6)	(5,5)	(6,4)	(7,3)	(6,2)	(7,2)	(8,4)	CO2	PO2	<b>8</b>
A	B	C	D	E	F	G	H														
(3,7)	(4,6)	(5,5)	(6,4)	(7,3)	(6,2)	(7,2)	(8,4)														
	c)	Employ the Apriori algorithm to derive association rules from the provided dataset by assuming a minimum support of 40% and a minimum confidence of 70%. 1 -- {Orange, Mango, Apple, Banana} 2 -- {Grapes, Kiwi, Mango} 3 -- {Apple, Banana} 4 -- {Grapes, Mango} 5 -- {Apple, Banana, Orange}	CO2	PO3	<b>8</b>																

\*\*\*\*\*