U.S.N.										

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: Computer Science and Engineering **Duration: 3 hrs.**
Course Code: 22CS6PCMAL **Max Marks: 100**
Course: 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.

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)	List and explain the different types of Machine learning with an example for each.	CO1	PO1	8
	b)	Mention the advantages of writing a function for data preparation than doing manual data preparation.	CO1	PO1	6
	c)	In general, it is easy to get a large amount of data for training, but it is not perfectly representative of the data that will be used in production. Analyze the situation and suggest the solution for the same. Justify your answer.	CO2	PO2	6
		OR			
2	a)	List and explain the main challenges of Machine Learning.	CO1	PO1	10
	b)	Assume that you now have a shortlist of promising models. You now need to fine-tune them. List and explain the ways in which you can fine tune the model and explain.	CO1	PO1	10
		UNIT - II			
3	a)	Define the following: i)Information gain ii)Entropy	CO1	PO1	2
	b)	Consider the below dataset and find in which class of people Angelina will lie whose k factor is 3 and age is 5.	CO2	PO2	8

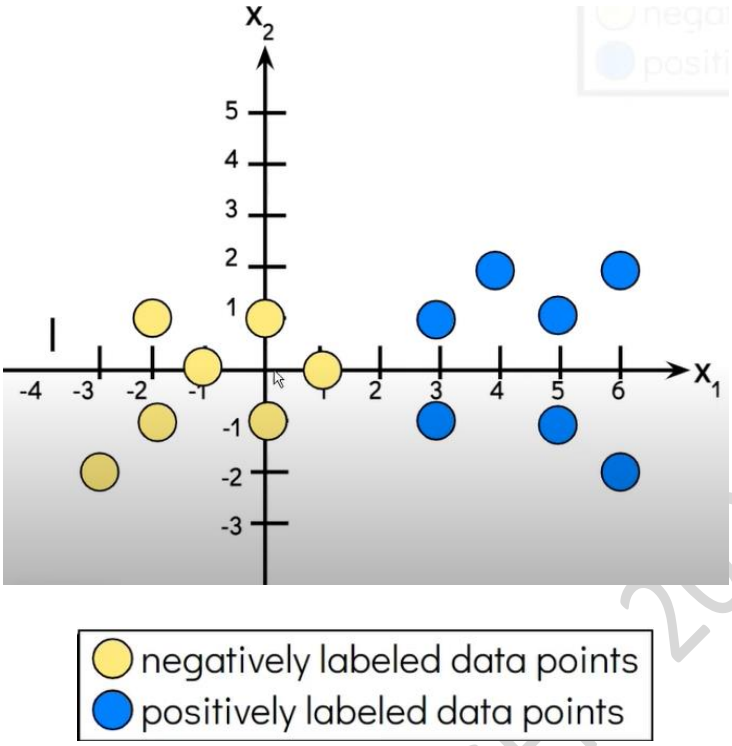
NAME	AGE	GENDER	CLASS OF SPORTS
Ajay	32	0	Football
Mark	40	0	Neither
Sara	16	1	Cricket
Zaira	34	1	Cricket
Sachin	55	0	Neither
Rahul	40	0	Cricket
Pooja	20	1	Neither
Smith	15	0	Cricket
Laxmi	55	1	Football
Michael	15	0	Football

	c)	Construct a Regression tree using the following data which consists of 10 data instances and three attributes. Result is the target attribute.	***CO3***	***PO3***	**10**

S.No.	Assessment	Assignment	Project	Result (%)
1.	Good	Yes	Yes	95
2.	Average	Yes	No	70
3.	Good	No	Yes	75
4.	Poor	No	No	45
5.	Good	Yes	Yes	98
6.	Average	No	Yes	80
7.	Good	No	No	75
8.	Poor	Yes	Yes	65
9.	Average	No	No	58
10.	Good	Yes	Yes	89

		OR			
4	a)	"KNN is called a lazy learning algorithm". Explain your answer on the statement.	***CO1***	***PO1***	**2**
	b)	Consider the below dataset, apply multiple linear regression for the below table using matrix method.	***CO2***	***PO2***	**8**

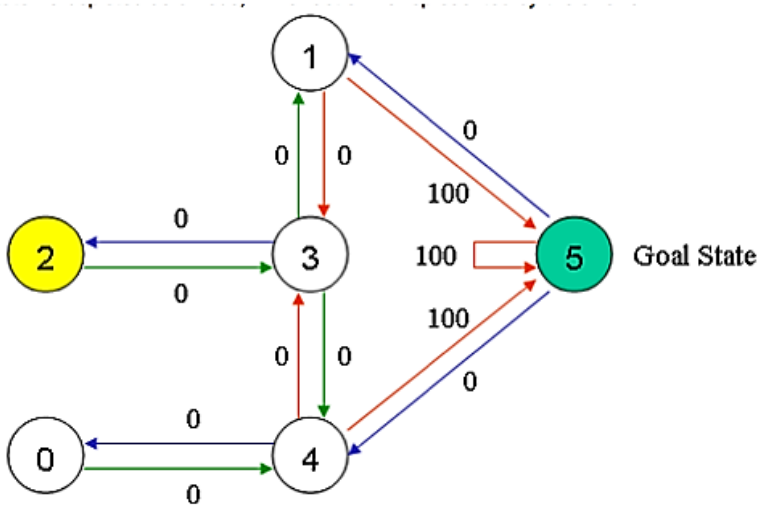
y	x1	x2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

		c)	Consider the below graph, Identify the support vectors and draw the hyperplane using Support vector machine.  	*CO2*	*PO2*	**10**
			UNIT - III			
	5	a)	List and explain the types of cross validation techniques in Machine learning.	*CO1*	*PO1*	**5**
		b)	Consider the confusion matrix of iris data set, find the following: i)Accuracy ii)Precision iii)Recall iv)F1 score of each class v) Macro F1 score vi) Weighted F1-score Predicted values Actual values Setosa: 7, 0, 0 Versicolor: 0, 10, 2 Virginica: 0, 2, 9	*CO2*	*PO2*	**10**
		c)	Identify the appropriate testing method used when we want to compare two learning algorithms and test whether they construct classifiers that have the same expected error rate and explain the same.	*CO2*	*PO2*	**5**
			OR			
	6	a)	Explain Bootstrap sampling method.	*CO1*	*PO1*	**5**

<table>
<tr><td></td><td></td><td>b)</td><td>List and explain the guidelines to be followed for Machine Learning experiments.</td><td>C01</td><td>PO1</td><td>10</td></tr>
<tr><td></td><td></td><td>c)</td><td>Consider that we have a training set and a validation set, we use two algorithms to train two classifiers on the training set and test them on the validation set contingency table and compute their errors. Consider the below contingency table and use the appropriate test to calculate the test statistic.

<table><tr><td>80</td><td>100</td></tr><tr><td>10</td><td>120</td></tr></table></td><td>CO2</td><td>PO2</td><td>5</td></tr>
<tr><td></td><td></td><td></td><td>UNIT_IV</td><td></td><td></td><td></td></tr>
<tr><td>7</td><td></td><td>a)</td><td>Write the k-means clustering algorithm by mentioning all the steps.</td><td>CO1</td><td>PO1</td><td>5</td></tr>
<tr><td></td><td></td><td>b)</td><td>Apply the Adaboost algorithm for the following dataset and classify the dataset with illness as target attribute.
Use 3 decision stumps for each of the three attributes

<table><tr><td>Gender</td><td>Age</td><td>income</td><td>Illness</td></tr><tr><td>Male</td><td>41</td><td>40000</td><td>Yes</td></tr><tr><td>Male</td><td>54</td><td>30000</td><td>No</td></tr><tr><td>Female</td><td>42</td><td>25000</td><td>No</td></tr><tr><td>Female</td><td>40</td><td>60000</td><td>Yes</td></tr><tr><td>Male</td><td>46</td><td>50000</td><td>Yes</td></tr></table></td><td>CO2</td><td>PO2</td><td>10</td></tr>
<tr><td></td><td></td><td>c)</td><td>Identify the ensemble techniques which trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output. List and explain the different types the above method uses for the classification technique.</td><td>CO2</td><td>PO2</td><td>5</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td>8</td><td></td><td>a)</td><td>Explain the Stacking ensembled method with a neat diagram.</td><td>CO1</td><td>PO1</td><td>5</td></tr>
<tr><td></td><td></td><td>b)</td><td>Apply K(=2)-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose c1 = (185, 72) and c2 = (170, 56) as initial centroids.</td><td>CO1</td><td>PO1</td><td>10</td></tr>
<tr><td></td><td></td><td>c)</td><td>Consider the two dimensional patterns (4,11),(8,4), (13,5),(7,14). Compute the covariance matrix for the given data set.</td><td>CO1</td><td>PO1</td><td>5</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>Explain dynamic Programming with key methods involved in it.</td><td>CO1</td><td>PO1</td><td>5</td></tr>
</table>

		b)	 Consider the above diagram, apply the Q-learning algorithm to find the optimal path from all the states to a goal state. Consider γ (gamma)value as 0.8.	***CO1***	***PO1***	**10**
		c)	Analyze a situation where you need the agent to have control over its actions and be able to explore different options to find the best solution. Identify which type of reinforcement method to be used for the above scenario. Explain the same.	***CO2***	***PO2***	***5***
			OR			
	10	a)	Explain Marcov decision process with a neat diagram and example?	***CO1***	***PO1***	***6***
		b)	Imagine you're planning a road trip from Banglore to Delhi, passing through Hyderabad place and Pachmarhi place. You want to estimate the total travel time for this trip. Each trip is an episode, and the travel times can vary due to factors like traffic and weather. The episodes of the trip are • **Trip 1**: Banglore -> Hyderabad (2 hours), Hyderabad -> Pachmarhi(3 hours), Pachmarhi -> Delhi (4 hours); Total: 9 hours • **Trip 2**: Banglore -> Hyderabad (1.5 hours), Hyderabad -> Pachmarhi (2.5 hours), Pachmarhi -> Delhi (4.5 hours); Total: 8.5 hours • **Trip 3**: Banglore -> Hyderabad (2 hours), Hyderabad -> Pachmarhi (3 hours), Pachmarhi -> Delhi(4 hours); Total: 9 hours Explain how Temporal difference method is more efficient than monto carlo method?	***CO2***	***PO2***	***8***
		c)	List and explain the several ways in which Generalization can be achieved in Reinforcement Learning.	***CO1***	***PO1***	***6***

******