U.S.N.										

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

Autonomous Institute Affiliated to VTU

October 2024 Supplementary 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

			UNIT - I	CO	PO	Marks
1	a)		Describe various key purposes of Machine learning.	CO1	PO1	05
	b)		Identify the type of algorithm to be used for the following scenario: i) Classification of email as spam or not ii) Photo-hosting services, such as Google Photos iii) Use to allow a robot to walk in various unknown terrains iv) You have a lot of data about your blog's visitors v) Algorithm relies on a similarity measure to make predictions	CO2	PO2	05
	c)		Write brief note on: i) Preparing a dataset to perform a Machine learning task ii) Online learning	CO1	PO1	10
			UNIT - II			
2	a)		For the transactions shown in the table below, compute the following: i) Entropy of the collection of transaction records of the table with respect to classification. ii) What is the information gain of a1 and a2 relative to the transactions of the table?	CO2	PO2	07
	b)		Describe the significance of Kernel functions in SVM. List any two kernel functions.	CO1	PO1	05

Instance	1	2	3	4	5	6	7	8	9
a1	T	T	T	F	F	F	F	T	F
a2	T	T	F	F	T	T	F	F	T
Target class	+	+	-	+	-	-	-	+	-

		c)	Based on a survey conducted in an institution, students are classified based on the two attributes of academic excellence and other activities. Given the following data, identify the classification of a student with X = 5 and Y = 7 using k-NN algorithm (choose k as 3). **X (Academic Excellence) | Y (Other Activities) | Z (Classification)** 8 | 6 | Outstanding 5 | 6 | Good 7 | 3 | Good 6 | 9 | Outstanding	*CO2*	*PO2*	**08**
			OR			
	3	a)	Given the set of values X = $(3, 9, 11, 5, 2)^T$ and Y = $(1, 8, 11, 4, 3)^T$. Evaluate the regression coefficients.	*CO2*	*PO2*	**07**
		b)	Using the data in table given below, construct a tree to predict the values of y. x_1: 1, 3, 4, 6, 10, 15, 2, 7, 16, 0 x_2: 12, 23, 21, 10, 27, 23, 35, 12, 27, 17 y: 10.1, 15.3, 11.5, 13.9, 17.8, 23.1, 12.7, 43.0, 17.6, 14.9	*CO3*	*PO3*	**08**
		c)	Consider the dataset with six data points: {(x1,y1),(x2,y2),…,(x6,y6)} , where $x_1 = \begin{bmatrix}1\\0\end{bmatrix}, x_2 = \begin{bmatrix}0\\1\end{bmatrix}, x_3 = \begin{bmatrix}0\\-1\end{bmatrix}, x_4 = \begin{bmatrix}-1\\0\end{bmatrix}, x_5 = \begin{bmatrix}2\\2\end{bmatrix}, x_6 = \begin{bmatrix}-2\\-2\end{bmatrix}$ and the labels are given by $y_1=y_2=y_5=1$, and $y_3=y_4=y_6=-1$. A hard margin linear support vector machine is trained on the above dataset. Identify the support vectors and draw the hyperplane.	*CO2*	*PO2*	**05**
			UNIT - III			
	4	a)	Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data.	*CO3*	*PO3*	**08**
		b)	Explain the necessity of combining several algorithms for accomplishing a particular task.	*CO1*	*PO1*	**06**
		c)	Explain: i) t-test ii) McNemar's test	*CO1*	*PO1*	**06**

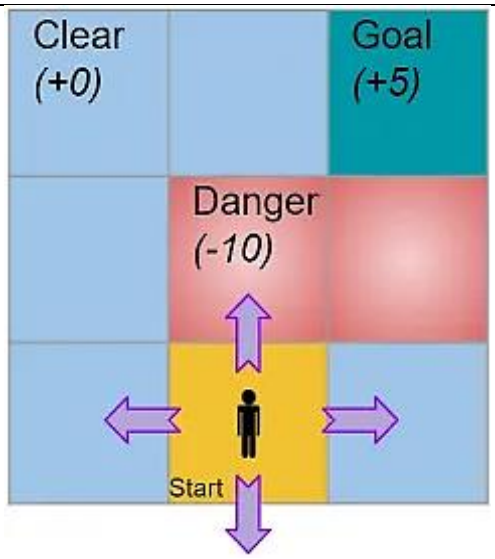
			UNIT – IV			
	5	a)	Apply the K-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to **two iterations** and show the clusters. Initially choose the **first two objects** as initial centroids.	***CO2***	***PO2***	**12**
		b)	Consider a training dataset of six data instances shown in the table. Use 4 Decision stumps for each of the 4 attributes. Apply AdaBoost algorithm and classify the dataset with Job Offer as target attribute.	***CO2***	***PO2***	**08**

CGPA	Interactiveness	Practical Knowledge	Job Profile
>=9	Yes	Good	Yes
<9	No	Good	Yes
>=9	No	Average	No
<9	No	Average	No

			OR			
	6	a)	Given the data in Table, reduce the dimension from 2 to 1 using the Principal Component Analysis (PCA) algorithm.	***CO2***	***PO2***	**12**

Feature	Example 1	Example 2	Example 3	Example 4
X_1	4	8	13	7
X_2	11	4	5	14

		b)	Use K Means clustering to cluster the following data into two groups. Assume cluster centroids are m1=2 and m2=4. The distance function used is Euclidean distance. { 2, 4, 10, 12, 3, 20, 30, 11, 25 }	***CO2***	***PO2***	**08**
			UNIT – V			
	7	a)	Consider a 3x3 grid, where the player starts in the Start square and wants to reach the Goal square as their final destination, where they get a reward of 5 points. Some squares are Clear while some contain Danger, with rewards of 0 points and -10 points respectively. In any square, the player can take four possible actions to move Left, Right, Up, or Down. Apply the Q-learning algorithm to find the optimal path from all the states to a goal state. Learning rate = 0.8	***CO2***	***PO2***	**12**

						
		b)	Differentiate between i)Positive Reinforcement and Negative Reinforcement ii)Dynamic Programming and Monte carlo method	*CO1*	*PO1*	**08**

******