Final Assessment Test – May 2024

VIT
Vellore Institute of Technology

Course: **PMCA507L - Machine Learning**

Class NBR(s): **3555/3560/3562**   Slot: **A2+TA2**

Time: **Three Hours**   Max. Marks: **100**

➤ KEEPING MOBILE PHONE/ELECTRONIC DEVICES EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE
➤ DON'T WRITE ANYTHING ON THE QUESTION PAPER

Answer any TEN Questions
(10 X 10 = 100 Marks)

1. Elucidate the concept of Machine Learning and brief about various types of ML algorithm and their application.

2. Make use of Candidate Elimination algorithm for learning a hypothesis space for a concept defined by the following training examples:

Color	Size	Shape	Texture	Edible
Green	Large	Round	Rough	Yes
Red	Small	Round	Smooth	No
Green	Medium	Oval	Rough	No
Red	Small	Oval	Smooth	No
Blue	Large	Round	Smooth	Yes
Red	Medium	Oval	Rough	No

3. Consider a dataset that represents the relationship between the size of a house (in square feet) and its selling price (in thousands):

House Size (x)	Selling Price (y)
1000	250
1500	350
2000	480
2500	520
3000	600

By building a simple linear regression model predict the selling price based of the house if size is 3500 and 4500. Write the regression line equation and plot the best fit line. Evaluate the mean square error and root mean square error.

4. Estimate the total error encountered in the following Neural Network. Use Sigmoid activation function.

Target Value
T1=0.01
T2=0.99
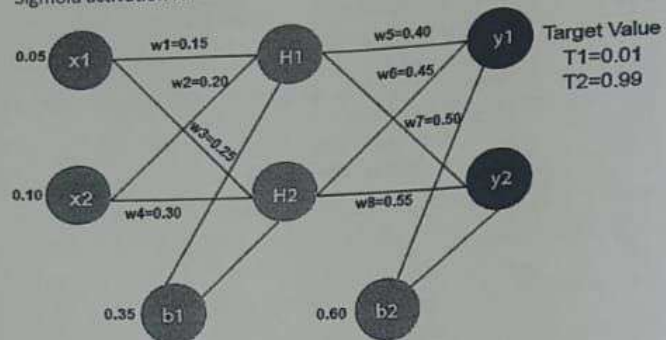
5. Consider a dataset representing the characteristics of fruits and whether they are "Sweet" or "Sour":

Color	Diameter (cm)	Weight (g)	Label
Red	7	150	Sweet
Yellow	6	130	Sour
Red	7	150	Sweet
Red	8	150	Sweet
Yellow	6	130	Sour
Green	8	180	Sour
Yellow	7	150	Sweet
Green	8	180	Sour
Green	8	180	Sour
Yellow	6	130	Sour

Build a decision tree by using gini index as Variable selection criterion to predict whether a fruit is "Sweet" or "Sour" based on the given features.

6. Consider a dataset representing the attributes of fruits and their classification as either "Apple" or "Orange": Develop a K-Nearest Neighbors (KNN) model to predict the classification of a fruit based on its weight, diameter, and color. With K value of 5, determine the classification of a fruit with the following attributes:

Weight: 155 grams Diameter: 7.3 cm Color: Red

Weight (grams)	Diameter (cm)	Color	Label
150	7	Red	Apple
170	7.5	Red	Apple
140	6.8	Orange	Orange
155	7.2	Red	Apple
145	6.9	Orange	Orange
160	7.4	Red	Apple
135	6.7	Orange	Orange

7. Enumerate in detail about graphical model with example.

   a) Brief about Ensembler learning with illustration. [2.5]

   b) Describe bagging with an example of your own. [2.5]

   c) Narrate the uses of stacked generalization. [2.5]

   d) Elucidate the voting concept and its type with neat sketch. [2.5]

8. Consider a dataset with two features, "Age" and "Income", and the target variable is "Purchased" which indicates whether a person purchased a product (1 for purchased, 0 for not purchased).

Age	Income	Purchased
20	50000	0
30	80000	0
25	60000	1
35	120000	1
45	90000	0

Make use of AdaBoost algorithm to classify whether a person with "Age = 28" and "Income = 70000" will make a purchase or not.

9. Apply K-Means Clustering algorithm to cluster the given points into 2 clusters. Find the final centroid point. A(1,1) B(1,2) C(2,1) D(4,3) E(5,4) F(4,5)

10. Consider the following dataset with two features, "Height" and "Weight" and Use the Hierarchical Clustering algorithm to cluster the given persons based on their height and weight into 3 clusters.

Person	Height(cm)	Weight(kg)
John	170	68
Alice	165	55
Bob	180	85
Carol	155	50
David	182	90
Eva	160	53
Frank	175	80

11. Analyse the following dataset with three features, "X", "Y", and "Z": Given data = { 2,4,5,7,1 ; 4,2,7,6,3; 1,3,6,8,2 }. Apply Principal Component Analysis (PCA) to reduce the dimensionality of the dataset to 2 principal components.

12. Imagine you are designing an online advertisement system. You have 3 different ads (arms) that you want to test to see which one yields the highest click-through rate (CTR). You can only show one ad to a user at a time, and after each display, you can either get a click (reward of 1) or no click (reward of 0). The goal is to maximize the total reward (total clicks).

Given the following reward history for each ad after 10 trials:

Ad A: 1, 0, 0, 1, 0, 1, 0, 0, 0, 1

Ad B: 0, 0, 1, 0, 0, 0, 1, 0, 1, 0

Ad C: 1, 0, 0, 0, 1, 0, 0, 0, 1, 0

Construct the ε-greedy strategy with $\varepsilon = 0.3$ to decide which ad to display to the user for the 11th trial. Break ties arbitrarily.

⇔⇔⇔ G/E/TX⇔⇔⇔