G/E/TX

Reg. No:

VIT
Vellore Institute of Technology

Final Assessment Test – May 2024

Course: **BITE410L - Machine Learning**

Class NBR(s): **3823/3974**

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. Critically differentiate the supervised, un-supervised, reinforcement machine learning techniques with respect to:

(i) Objective (ii) Data (iii) Feedback mechanism (iv) Use-cases

2. The following dataset contains the features of eight samples of two discriminant classes namely A, and B. Compute Within-Class Scatter Matrix for the above classes as part of performing Linear Discriminant Analysis.

Sample	Feature 1	Feature 2	Class
1	2	3.5	A
2	1.5	2.5	A
3	3	4	A
4	2.5	3.8	A
5	3.5	2	B
6	4	2.7	B
7	5	3	B
8	4.5	2.5	B

3. The amount of money spent on R &D in crores (X), and Sales data in crores ('Y') are given in the table. Apply linear regression technique to predict the sales when the money spent on R & D is Rs.9 crores. Draw the regression line for the given data.

X (Money spent on R & D in crores)	Y (Sales data in crores)
4	**36.3**
6	**40.1**
6	**41.2**
8	**53.0**
6	**44.0**
7	**45.6**
5	**39.0**
7	**47.5**

4. Consider a training data set consisting of the fauna of the world. Each unit has three features namely "Swim", "Fly" and "Crawl". Let the possible values of these features be as follows:

Swim - Fast, Slow, No

Fly - Long, Short, Rarely, No

Crawl - Yes, No

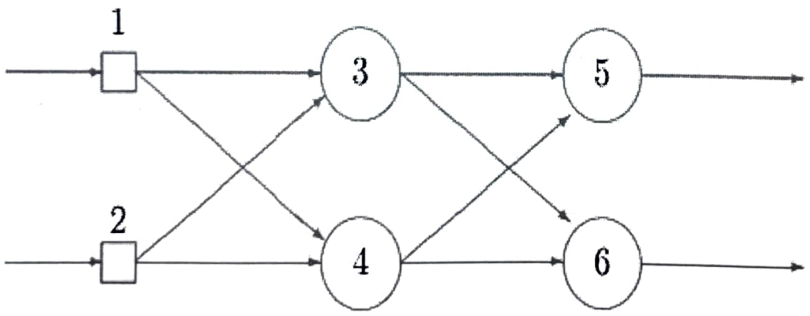
For simplicity, each unit is classified as "Animal", "Bird" or "Fish". Let the training data set be as in the following table. Use Naive Bayes algorithm to classify a particular species if its features are Slow, Rarely, No.

Sl. No.	Swim	Fly	Crawl	Class
1	Fast	No	No	Fish.
2	Fast	No	Yes	Animal
3	Slow	No	No	Animal
4	Fast	No	No	Animal
5	No	Short	No	Bird .
6	No	Short	No	Bird.
7	No	Rarely	No	Animal
8	Slow	No	Yes	Animal
9	Slow	No	No	Fish.
10	Slow	No	Yes	Fish
11	No	Long	No	Bird.
12	Fast	No	No	Bird

5. The points (4, 1), (4, -1) and (6, 0) belong to 'positive' classes and the points (1, 0), (0, 1) and (0, -1) belong to 'negative' classes. In order to classify the new points (5, 2), (7, 2) and (-1, -2), construct an optimal hyperplane using the linear SVM classifier.

X1	X2	Class
4	1	Yes
4	-1	Yes
6	0	Yes
1	0	No
0	1	No
0	-1	No

6. The following diagram represents a feed-forward neural network with one hidden layer:



A weight on connection between nodes i and j is denoted by w_{ij}, such as w_{13} is the weight on the connection between nodes 1 and 3. The following table lists all the weights in the network:

$w_{13} = -2$	$w_{35} = 1$
$w_{23} = 3$	$w_{45} = -1$
$w_{14} = 4$	$w_{36} = -1$
$w_{24} = -1$	$w_{46} = 1$

Each of the nodes 3, 4, 5 and 6 uses the following activation function:

$$\varphi(v) = \begin{cases} 1 \text{ if } v \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

where v denotes the weighted sum of a node. Each of the input nodes (1 and 2) can only receive binary values (either 0 or 1). Calculate the output of the network (y_5 and y_6) for each of the input patterns:

Pattern	P1	P2	P3	P4
Node 1:	0	1	0	1
Node 2:	0	0	1	1

7. The following dataset has the information about hair color, eye color, and skin color of persons. Hair color, eye color, and skin color are all categorical variables. Explain the steps involved in K-Mode clustering algorithm. Choose P1, P7, P8 as leaders/clusters. Create three clusters from the following data set using the K-Mode clustering algorithm with Hamming distance.

person	hair color	eye color	skin color
P1	blonde	amber	fair
P2	brunette	gray	brown
P3	red	green	brown
P4	black	hazel	brown
P5	brunette	amber	fair
P6	black	gray	brown
P7	red	green	fair
P8	black	hazel	fair

8. Explain the steps involved in bagging and boosting with a neat diagram and also differentiate bagging and boosting in machine learning.

9. The following confusion matrix resulted in the classification of the IRIS dataset into three classes namely Setosa, Virginica, and Versicolor by a bench-marking supervised machine learning algorithm for a sample data set. Compute the performance of the learning algorithm by calculating the Accuracy, Error, Precision, Recall, F1-score for each predicted class label.

		Predicted class		
		Setosa	Versicolor	Virginica
Actual class	Setosa	10	0	0
	Versicolor	0	17	1
	Virginica	0	0	11

10. With a neat sketch, discuss the steps involved in Q-learning, as an iterative technique in a reinforcement learning context.

11. Examine the following table, which displays eight data points with two dimensions (A and B) that indicate position. Assuming the following starting seed values: Centriod1 (Y1) = (1,1), Centriod2(Y4) = (5,7), and Centriod3(Y7) = (1,2). Between two points, use the Euclidean distance. Apply the k means algorithm to identify three clusters.

	A	B
Y1	1	1
Y2	6	7
Y3	4	6
Y4	5	7
Y5	5	2
Y6	2	3
Y7	1	2
Y8	3	1

12. With a neat sketch, discuss the various sub-elements of a reinforcement learning system with an illustrative example.

⇔⇔⇔ G/E/TX ⇔⇔⇔