U/D/TY

Reg. No: 22BIT0601

VIT
Vellore Institute of Technology

Final Assessment Test – April 2025

Course: BITE410L - Machine Learning

Slot: A1+TA1

Class NBR(s): 3062/3065/3068

Max. Marks: 100

Time: Three Hours

- KEEPING MOBILE PHONE/ANY ELECTRONIC GADGETS, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE
- DON'T WRITE ANYTHING ON THE QUESTION PAPER

Answer ALL Questions

(10 X 10 = 100 Marks)

1. Explain the meaning of 'Well-posed learning problem' in the context of machine learning. How it differs from an ill-posed-learning? With your own intuition, give two examples for well-posed learning with deliberate explanation of their necessary components.

2. Suppose you have a multi-class classification problem with two categorical input features X1 and X2, where X1 can take 3 values {A,B,C}, X2 can take 2 values {D,E}, and the dependent variable Y can take four possible values {0,1,2,3}. Compute the total number of hypothesis that a learner would learn from the possible mapping the input space to the output space. If the learning algorithm is considered as PAC learnable, compute the minimum number of training examples needed to achieve a high probability of success in generalizing the unseen instances. The approximation error parameter,'ε' and the probability of failure,'δ 'could be considered as 0.10, and 0.05 respectively.

3. With a neat diagram, discuss the three broad taxonomy of feature selection approaches used in machine learning for the task of data pre-processing. Illustrate the implication of various statistical tests, and techniques used in each approach.

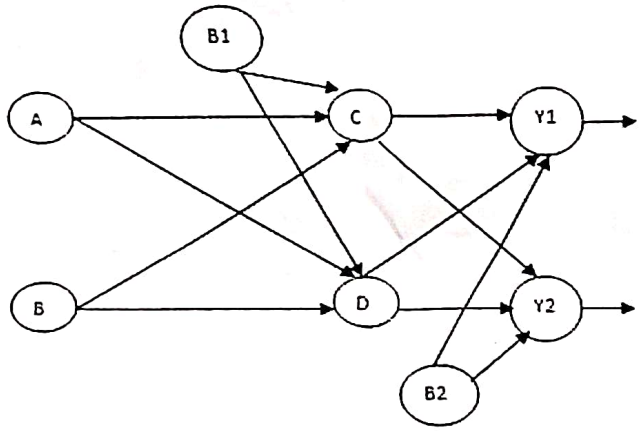
4. Perform Principal Component Analysis for the following features and compute the first principal component (PC_1) which explains the most data variance.

Feature	Example 1	Example 2	Example 3	Example 4
A	2	1	0	-1
B	4	3	1	0.5

5. Given the following data on a certain set of patients seen by a doctor, using naive Bayes classification algorithm, can the doctor conclude that a person having chills, fever, mild headache and without running nose has the flu? Calculate the posterior probability for the test sample.

Chills	Running Nose	Headache	Fever	Has Flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

6. *The following* Artificial Neural Network has been constructed for the task of regression. The input samples at the neurons A, and B are 0.05, 0.10 respectively. A bias weight of 0.35, and 0.60 is applicable for the hidden layer, and output layer respectively. The arbitrarily assigned weights are as follows :



W_{AC}	W_{AD}	W_{BC}	W_{BD}	W_{CY1}	W_{CY2}	W_{DY1}	W_{DY2}
0.15	0.20	0.25	0.30	0.40	0.50	0.45	0.55

The sigmoid activation function is used on both hidden and output layers. The output layer has two output namely, neuron Y1, and neuron Y2 with the output values of 0.01, and 0.99 respectively. The difference between actual and predicted values can be calculated using the loss function: $\frac{1}{2}(t-y)^2$.

Compute the gradients of the total loss, perform back propagation using Stochastic Gradient Descent (SGD) optimization technique, and calculate the new weight of W_{CY1}. Use a learning rate of 0.50 while applying weight updation rule.

7. The following dataset contains the training instances of Class A, and Class B with respect to the independent features x1, and x2.

x1	x2	Class Label
2	2	A
-2	2	A
-2	-2	A
2	-2	A
4	0	B
0	4	B
-4	0	B
0	-4	B

Determine the weight vector and bias for the hyper plane that accurately discriminate the two classes using non-linear SVM technique with the following mapping function :

$$\emptyset\binom{x1}{x2} = \begin{cases} 8 - x1 + (x1 - x2)^2 & \text{if } \sqrt{x1^2 + x2^2} \geq 4 \\ 8 - x1 + (x1 - x2)^2 & \end{cases}$$

$\binom{x1}{x2}$ Otherwise.

8.(a) Suppose that your machine learning task is to cluster the following instances into two distinct clusters using K-means clustering algorithm, perform three iterations and show how objects of the clusters have been iteratively relocated at the end of each iteration. Initially choose instance 1 as the centroid of the first cluster, and choose instance 3 as the centroid of the second cluster. Use Euclidian distance as the distance measure. Tabulate the results of each iteration, which include new centroids of the clusters, and objects belong to respective clusters.

Instance	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

OR

8.(b) Explain the detailed steps involved in the agglomerative technique based hierarchical clustering algorithm using single-link approach for the below data points and draw the final dendrogram to show the formation of clusters. Use Euclidian distance as the distance measure for the formation of proximity matrix.

Points	P1	P2
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.31

9. With neat sketch, illustrate the functionalities of various boosting algorithms used in sequential ensemble models. Critically compare the Pros, and Cons of each algorithm in overcoming the limitations of single-hypothesis machine learning model.

10.a) With schematic illustration, discuss any two novel applications of reinforcement learning and explain why such applications cannot be dealt with the Supervised & Un-supervised learning counterparts.

OR

10.b) Justify why Q-Leaning is called as off-policy learning algorithm in a reinforcement learning context. With an example, illustrate how Bellman optimality equation describes the optimal state-value function for episodic task involved in Q-Learning.

⇔⇔⇔ **U/D/TY** ⇔⇔⇔