


VIT

Vellore Institute of Technology

Final Assessment Test – November 2025

Course: BITE410L - Machine Learning

Class NBR(s): 3194/3203/7351

Time: Three Hours

Slot: C2+TC2

Max. Marks: 100

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

COs	CO Statements
CO1	Understand the various Machine learning approaches and data preprocessing techniques
CO2	Develop the learning models for the problems using supervised and advanced supervised learning-based approaches
CO3	Apply the different clustering approaches to handle unsupervised based learning
CO4	Examine the various ensemble models
CO5	Infer the reinforcement learning

BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)

Answer ALL Questions
(10 X 10 = 100 Marks)

1. Explain unsupervised learning through an architectural diagram. Compare supervised learning with unsupervised learning with respect to the parameters; Definition/ Learning policy, Algorithms applied/ Process, Goal, Feedback and Example Applications. CO1 BL1

2. Determine the Pearson's correlation coefficient for the tabular data given below. CO1 BL3

Kgs of pesticides used per acre	Production in Kgs	Kgs of pesticides used per acre	Production in Kgs
10	25	60	67
20	37	70	73
30	45	80	79
40	52	90	85
50	61	100	92

3. Use Linear Discriminant Analysis method to find the eigen vector corresponding to the largest eigen value for the two sets of data given below: CO1 BL4

$$W_1 = \{(2,1), (1,2), (3,2), (2,3), (1,4)\} \text{ and}$$

$$W_2 = \{(5,4), (6,5), (7,4), (8,5), (9,6)\}.$$

4. Using a Gaussian Naive Bayes classifier, how can we predict whether an employee will stay or leave the company based on their weekly work hours and stress level using the continuous data provided in the table below?

CO2 BL4

Work Hours	Stress Level	Attrition
40	3	Stay
55	8	Leave
38	2	Stay
60	9	Leave
45	5	Stay
50	7	Leave
42	4	Stay
58	8	Leave

Predict attrition for an employee working 48 hours/week with a stress level of 6.

5. Taking two input nodes and one output node Implement XOR logic function using M-P neuron with an intermediate layer containing two nodes.

CO2 BL3

6. Consider the customer dataset given below.

CO3 BL5

Customer ID	Favourite Category	Payment Method	Shopping Time
C1	Electronics	Credit Card	Evening
C2	Clothing	Cash	Morning
C3	Electronics	Cash	Evening
C4	Groceries	Credit Card	Afternoon
C5	Clothing	Credit Card	Morning
C6	Groceries	Cash	Afternoon
C7	Electronics	Credit Card	Morning
C8	Clothing	Cash	Evening
C9	Groceries	Credit Card	Morning
C10	Electronics	Cash	Afternoon

Use K-mode clustering algorithm to form 3 clusters from the data points. Take C1, C4 and C9 as the initial centroids.

- 7/ Classify the point $E = (3, 4)$ using the KNN algorithm with $K = 3$ using the following data table:

CO3 BL5

Point	X	Y	Class
A	1	2	Red
B	2	3	Red
C	3	3	Blue
D	6	5	Blue
E	4	2	Red
F	5	4	Blue
G	3	5	Blue

8. What is XGBoost algorithm? Explain its key features. Discuss on how it works and explain the basic steps.

CO4 BL2

- 9.a) Define Bias and variance in a Machine Learning model. What is meant by bias-variance trade off? State and explain the strategies followed to overcome bias-variance trade off.

CO4 BL2

OR

- 9.b) Define Bagging as used under Ensemble learning. Explain the algorithms based on bagging. Draw a comparison between bagging and boosting techniques.

CO4 BL2

- 10.a) What is reinforcement learning (RL)? Mention its basics and its architecture. What is the role of an agent in the context of RL? Define a framework in RL and explain its components.

CO5 BL1

OR

- 10.b) Define Markov Decision process. Explain its components elaborately. Explain the relevance of Markov Decision process to reinforcement learning.

CO5 BL1

⇔⇔⇔ X/K/TY ⇔⇔⇔