



VIT
Vellore Institute of Technology

Final Assessment Test – November 2024

Course: BCSE209L - Machine Learning

Class NBR(s): 1586/1592/1597/1685/1693/1695/1705
/1711/1717/1724/1728/1735

Slot: B1+TB1

Time: Three Hours

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

Answer ALL Questions

(10 X 10 = 100 Marks)

1. A real estate agency wants to develop a model to predict house prices based on various features such as location, size, number of bedrooms, and age of the property. How can machine learning be applied to this prediction task? Additionally, describe the data preprocessing steps necessary before using a machine learning algorithm.
2. A car dealership wants to predict the price of a car based on its age (in years). The following data is available:

Age (Years)	Price (\$)
1	10,000
2	25,000
3	20,000
4	15,000
5	12,000
6	17,000
7	21,000
8	26,000

- a) What is the equation of the regression line?
- b) Predict the price of a car that is 4.5 years old.

3. A multivariate logistic regression model has been built to predict the propensity of shoppers to perform a repeat purchase of a free gift that they are given. The descriptive features used by the model are the age of the customer, the average amount of money the customer spends on each visit to the shop, and the average number of visits the customer makes to the shop per week. The marketing department is using this model to determine who should be given the free gift. The weights in the trained model are shown in the table below

Feature	Weight
Intercept ($w[0]$)	-1.9112
Age	-0.0149
Shop Frequency	0.3728
Shop value	0.0149

Use this model to make predictions for each of the following query instances.

ID	Age	Shop Frequency	Shop value
1	46	0.8	54.66
2	25	2.46	5.64
3	38	0.6	80.56
4	25	0.36	85.32
5	35	0.54	82.69

- Describe the key steps involved in the XGBoost algorithm, starting from data input to model output. How does it differ from the Ada boosting algorithm?
- You are tasked with predicting customer churn for a subscription service. Your dataset shows that only 10% of customers churn while 90% do not. How would you apply SMOTE to address this imbalance, and what metrics would you use to evaluate your model's performance after applying SMOTE?
- Apply the candidate elimination algorithm to obtain the final version space for the following training sample.

ID	Animal	Color	Size	Habitat	Domestic
1	Dog	Brown	Medium	Urban	Yes
2	Dog	Brown	Small	Urban	Yes
3	Elephant	Brown	Big	Forest	No
4	Lizard	Green	Small	Forest	No
5	Cat	Black	Medium	Urban	Yes

- Consider the following data samples for a two-class problem:

Positive samples: $\{(1,2), (2,3), (3,0)\}$

Negative samples: $\{(5,1), (5,-1), (6,1)\}$

Draw the optimal hyperplane using a support vector machine.

- Consider the following dataset:

Age group	Cholesterol	Blood pressure	Diabetes	Risk level
Young	High	High	Yes	High
Young	Low	Low	No	Low
Middle	High	High	No	High
Middle	Low	High	Yes	High
Old	Low	Low	No	Low
Old	Medium	Low	No	Low
Young	Low	Medium	Yes	High

Construct a decision tree using the ID3 algorithm.

- 9.a) Given the following dataset with 10 samples. Compute a partitioning using the K-Modes clustering algorithm. Use samples 1, 5, and 8 as the initial cluster centroids. Perform clustering for 2 iterations. At the end of each iteration, display the updated distance matrix and the newly constructed centroids.

Sample	Color	Shape	Size	Material
1	Red	Circle	Small	Plastic
2	Blue	Square	Large	Metal
3	Green	Triangle	Medium	Wood
4	Yellow	Circle	Small	Plastic
5	Blue	Square	Medium	Metal
6	Red	Triangle	Large	Wood
7	Green	Circle	Small	plastic
8	Yellow	Square	Large	Metal
9	Red	Triangle	Medium	Wood
10	Blue	Circle	Small	Metal

OR

- 9.b) Consider the following dataset:

Feature	Example 1	Example 2	Example 3	Example 4
X1	6	10	15	9
X2	13	6	7	16

Reduce the dimension from 2 to 1 using Principal Component Analysis (PCA).

- 10.a) Describe the fundamental components of the reinforcement learning framework. Discuss a few applications of reinforcement learning.

OR

- 10.b) Describe the key components of the Q-learning algorithm. Explain the role of each component in the learning process. Additionally, provide examples to illustrate the significance of these components in a typical Q-learning scenario.

⇔⇔⇔ B/L/TX ⇔⇔⇔