Program: MCA / MTECH **Session:** 2025-2026

Course Nam e: DATA Science

Note/Instruction (If any): Attempt all the questions

1. Explain the difference between supervised and unsupervised learning. Provide examples of algorithms for each type of learning and discuss scenarios where each would be appropriately used.
2. Describe the concept of hypothesis space in machine learning. Explain the algorithm Find-S. Find a Hypothesis by Find-S for the following instances in a training set.

Origin	Manufacturer	Color	Decade	Type	Class
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive

3. Explain the purpose of training, test, and validation sets in machine learning. What is cross-validation, and how does it differ from a simple train-validation-test split?
4. Given the confusion matrix, find: Classification Accuracy, Precision, Recall, Type 1 Error, Type 2 Error, F1-score, Sensitivity, Specificity. Explain the importance of Precision, Recall as compare to Accuracy with an example.

		Predicated	
		Positive	Negative
Actual	Positive	6	4
	Negative	2	8

5. Discuss the importance of dimensionality reduction in machine learning. Compare and contrast feature extraction and feature selection as two approaches to dimensionality reduction, and provide examples of techniques used for each approach.
6. Define overfitting and underfitting in the context of machine learning models. Describe how these issues can be identified and list strategies to prevent or mitigate them.
7. Differentiate logistic regression and linear regression on the basis of hypothesis, parameters and cost function. How Gradient descent algorithm is used to optimize the parameter?

8. Write a program in Python to implement logistic regression and linear regression. For every method compute and display the implemented model's Evaluation Measures (whichever is applicable) such as accuracy and other values of different evaluation metrics such as Confusion Matrix, Accuracy, Precision, Recall, F1 Score, Specificity and Loss Function, RMSE and R2 score.
9. Write one hot encoding for the given data.

Fruit	Value of Fruit	Price
Apple	1	5
Mango	2	10
Apple	1	15
Orange	3	10

10. Why K-NN is consider as lazy learning? Discuss k - Nearest Neighbors. Consider the following table – it consists of the height, age and weight (target) value for 10 people. The weight value of ID11 is missing. Predict the weight of this person (ID11) based on their height and age.

ID	Height	Age	Weight
1	5	45	77
2	5.11	26	47
3	5.6	30	55
4	5.9	34	59
5	4.8	40	72
6	5.8	36	60
7	5.3	19	40
8	5.8	28	60
9	5.5	23	45
10	5.6	32	58
11	5.5	38	?