

PSG COLLEGE OF TECHNOLOGY, COIMBATORE - 641 004

Department of Applied Mathematics and Computational Sciences

M. Sc TCS- Semester 5

CONTINUOUS ASSESSMENT TEST 1 Date: 28/8/2025

20XT53 - Machine Learning

Time: 1 Hour 30 min.

Maximum Marks: 40

**INSTRUCTIONS:**

1. Answer **ALL** questions. Each question carries 20 Marks.
2. Subdivisions (a)(i) and (a)(ii) carries 2 marks each, subdivision (b) carries 6 marks each and subdivision (c) carries 10 marks each.
3. Subdivisions (a) and (b) will be with no choice and Subdivision (c) may be with choice but not in more than 1 question.
4. \_\_\_\_\_ Data book / \_\_\_\_\_ table(s) may be permitted.
5. Course Outcome Table : 

Qn.1	CO1
------	-----

Qn.2	CO2
------	-----

- 1a Determine whether the function has a local maxima or minima at each of the critical points. L2  
 (i)  $f(x) = (x^2 - 1)^3$
- (II) Consider the weigh vector of a linear classifier for a binary classification problem whose labels lie in the set  $\{0,1\}$ . L2

$$\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

- If the bias of the classifier is set to 0, Write an equation of the decision boundary for this classifier?
- b Identify and examine the methods of outlier analysis in salary field of the given data points and predict the salary based on the experience with value of 9. Apply regression loss function to determine how well the algorithm fits the data. L3

Experience(X)	Salary(Y) (in lakh)
2	3
6	10
5	4
7	13

- c Explain in detail about the gradient descent algorithm. Apply gradient descent learning algorithm on the given data L4  
 $X = ([1, 2, 3, 4, 5])$   
 $Y = ([2, 4, 6, 8, 10])$   
 Initialize weights  $w = 0$  and  $b = 0$  and learning rate is 0.01. Use Mean square error has a cost function (J) and apply gradient learning and repeatedly adjust them in the direction of steepest descent. Update using the below formula for one iteration

$$w_j \leftarrow w_j - \alpha \frac{\partial J}{\partial w_j}$$

Discuss the about the effects of high and low learning rates.

2a(i) Consider learning three different classifiers C1, C2, C3 on a given data set such that C1 has high training as well as test accuracies, C2 has high training accuracy but low-test accuracy, whereas C3 has low training as well as test accuracies L3

Which one of the following statements is correct?

C2 is overfitting whereas C3 is underfitting.

C2 is underfitting whereas C3 is overfitting

Justify your answer

2a(ii) Suppose you are using polynomial regression, you plot the learning curves and you notice that there is a large gap between the training error and the validation error. Mention the ways to solve this problem. L2

b Apply Navie Bayesian classification on the given data L4

Slno	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

Identify the species of an entity with the following attributes.

$X = \{\text{Colour}=\text{Green}, \text{Legs}=2, \text{Height}=\text{Tall}, \text{Smelly}=\text{No}\}$

c Describe how an odds ratio is calculated and how it is interpreted in the context of logistic regression. How to perform a transformation of linear to logistic regression? Consider the following data and predict probabilities for a customer with an annual income of ₹50,000 with the threshold of 50% and above for purchase otherwise no purchase. The coefficient value for  $b_0 = -4$  and  $b_1 = 0.05$  L5

Customer	Annual Income (in ₹)	Purchased (1/0)
1	20,000	0
2	40,000	0
3	60,000	1
4	80,000	1