

COURSE CODE (CREDITS): 20B1WEC532 (3)

MAX. MARKS: 25

COURSE NAME: Introduction to Machine Learning

COURSE INSTRUCTORS: Lt. Pragya Gupta

MAX. TIME: 1 Hour 30 Minutes

Note: (a) All questions are compulsory.

(b) Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q1. Environmental scientists want to solve a two-class classification problem for predicting whether a population contains a specific genetic variant. They can use a confusion matrix to determine how many ways automated processes might confuse the machine learning classification model they're analyzing. Assuming the scientists use 500 samples for their data analysis, a table is constructed for their predictive and actual values before calculating the confusion matrix. After creating the matrix, the scientists analyze their sample data. Assume the scientists predict that 350 test samples contain the genetic variant and 150 samples don't. If they determine the actual number of samples containing the variant is 305, the actual number of samples without the variant is 195. These values become the "true" values in the matrix. Using the data from the confusion matrix, Calculate:

- a) Recall [6] (CO-4)
- b) Specificity
- c) Accuracy
- d) Miss classification or error rate

Q2. Given a Matrix A=

[4] (CO-1)

-1	1	2
3	-1	1
-1	3	4

Calculate and Show that $(A^{-1})^T = (A^T)^{-1}$

Q3. What is gradient descent algorithm in Machine Learning? Why it is used? [3] (CO-2)

Q4. What is Box Whisker Plot? How to create it? What are the uses of Box Whisker Plot? [3] (CO-2)

Q5. Find out the derivative of cost function (J) for the simple linear regression with one variable. [3] (CO-3)

Q6. Write a note on the following:

a) Bias- Variance Trade-off

b) Silhouette coefficient

[3+3] (CO-3)