



End Term (Odd) Semester Examination November 2025

Roll no. 236110242

Name of the Course and semester: B.TECH(CSE)-V SEMESTER

Name of the Paper: Machine Learning

Paper Code: TCS509

Time: 3 hour

Maximum Marks: 100

**Note:**

- (i) All the questions are compulsory.
- (ii) Answer any two sub questions from a, b and c in each main question.
- (iii) Total marks for each question is 20 (twenty).
- (iv) Each sub-question carries 10 marks.

Q1.

(2X10=20 Marks)

- Write a Python program that reads a text file, calculates how often each word appears, and saves the results in a dictionary arranged in order of word frequency. CO1
- Provide an overview of reinforcement learning. How does it differ from supervised and unsupervised learning paradigms? CO2
- Why is handling missing data a crucial step in data analysis? What are some common techniques for replacing values in a dataset during data transformation? CO2

Q2.

(2X10=20 Marks)

- Consider the data set shown in table below. Estimate all the parameters for training Naïve Bayes classifier and classify the given test sample ( $A=1, B=1$ ) using estimated parameters. CO3

A	B	Class
1	0	Positive
0	0	Negative
0	1	Negative
0	1	Negative
0	0	Positive
1	0	Positive

- Explain how Support Vector Machine can be used for classification of linearly separable data. CO3
- Explain the Curse of Dimensionality. How does it affect the performance of machine learning models, especially clustering? CO3

Q3.

(2X10=20 Marks)

- Consider the following training set having three different attributes namely  $x_1, x_2, x_3$ :

CO4

X1	X2	X3	Class
-1	1	1	Negative
5	1	0	Positive
0	2	1	Negative
-4	-1	0	Negative
0	0	1	Positive
9	2	1	Positive
2	2	0	Negative



**End Term (Odd) Semester Examination November 2025**

Using the K-nearest algorithm (kNN) algorithm with K=3, Classify the following points

- (i) (1,1,2)
- (ii) (2,1,1)
- (iii) (0,0,0)

b. Explain the logistic regression model. Derive the loss function using cross-entropy.

CO4

c. Given the covariance matrix:

CO4

$\Sigma =$	2	0.8
	0.8	0.6

- (i) Compute the eigenvalues and eigenvectors.
- (ii) Identify which principal component captures the maximum variance.

Q4.

(2X10=20 Marks)

a. Apply the K-means algorithm (one iteration only) on the following data points with k=2: Points: (1, 2), (1, 4), (5, 6), (6, 5). Initial centroids: (1, 2) and (5, 6).

CO5

b. What are the main differences and similarities between Matplotlib and Seaborn for data visualization? Suggest an appropriate Matplotlib or Seaborn plot to visualize the distribution of a single numerical variable. Justify your choice.

CO5

c. Given the points A(3, 7), B(4, 6), C(5, 5), D(6, 4), E(7, 3), F(6, 2), G(7, 2), H(8, 4), I(3,3), J(2,6), K(3,5), and L(2,4). Find the core points and outliers using DBSCAN. Take Eps = 1.9 and MinPts = 4. CO5

Q5.

(2X10=20 Marks)

a. Discuss different types of cross validation techniques in details.

CO6

b. Discuss various performance measures used by supervised and unsupervised machine learning models with their formula, preferences according to use and importance.

CO6

c. Write the short notes on: i) Hyperparameter Tuning ii) Random Search iii) Hierarchical Clustering. CO6

Note For the question paper setters:

- Question paper should cover all the COs of the course.
- Please specify COs against each question.