



Gandhi Institute of Engineering and Technology University, Odisha, Gunupur
(GIET University)

B. Tech. (Fourth Semester - Regular) Examinations, April 2026
23BCMPC24001 – Introduction to Machine Learning
(CSE AIML)

Time: 3 hrs

Maximum: 60 Marks

(The figures in the right hand margin indicate marks)

PART – A

Q.1. Answer ALL questions

a. How Machine Learning algorithms help in detecting fraudulent activities in Banking?
Also state some of the real-time software used.

b. Calculate Accuracy, Precision, Recall, and F1-score.

c. Explain the following regression performance metrics (MSE, RMSE, MAE and R²) in detail and discuss when each is preferred.

d. Explain the concepts of Bias, Variance, and the Bias–Variance Trade-off. Also include graphical interpretation.

e. Discuss the difference between Linear SVM Classification and Nonlinear SVM Classification with an example.

(2 x 5 = 10 Marks)

CO #

Blooms Level

CO1

K1

CO3

K2

CO4

K2

CO3

K1

CO6

K2

PART – B

Answer ALL the questions

2. a. Given the dataset, Compute the regression line Y=a+bX, Predict the value of Y when X = 5 and Calculate Mean Squared Error (MSE)?

b. Consider the following dataset
Let K = 2, Initial medoids are M1=A(2,6), M2=E(6,2), Assign points to the nearest medoid using Euclidean distance.

(10 x 5=50 Marks)

Marks

CO #

Blooms Level

5

CO2

K3

5

CO2

K3

Object	A	B	C	D	E	F	G
X	2	7	3	4	6	9	2
Y	6	2	5	5	2	1	9

(OR)

c. Consider the training dataset given in the following table. Use k-NN and determine the class. Test instance (7.6, 60, 8) and K=3.

d. A binary classifier produces the following predicted scores for 10 samples. Design the ROC curve.

3.a. Demonstrate the working of Naïve Bayes classifier Algorithm using a step-by-step numerical example?

b. Fit a second-degree polynomial regression model $y=ax^2+bx+c$ for the data.

5

CO1

K2

5

CO1

K3

5

CO2

K3

5

CO3

K2

S/No.	CGPA	Assessment	Project Submitted	Result
1	9.2	85	8	Pass
2	8	80	7	Pass
3	8.5	81	8	Pass
4	6	45	5	Fail
5	6.5	50	4	Fail
6	8.2	72	7	Pass
7	5.8	38	5	Fail
8	8.9	91	9	Pass

Sample	1	2	3	4	5	6	7	8	9	10
Actual Class	Positive (1)	Negative (0)	Positive (1)	Negative (0)	Positive (1)	Negative (0)	Positive (1)	Negative (0)	Positive (1)	Negative (0)
Predicted Score	0.9	0.95	0.8	0.75	0.7	0.65	0.6	0.55	0.5	0.45

x	1	2	3	4
y	2	7	6	9

Page 1 of 2

(OR)

- c. Explain cross-validation in machine learning. Explain the different types of cross-validations with an example? 5 CO2 K2
- d. What is KNN? Where are utilising the concept? List the importance with example? 5 CO4 K2
- 4.a. You are given a set of one-dimensional data points: {5, 10, 15, 20, 25, 30, 35}. Assume that $k = 2$ and first set of random centroid is selected as {15, 32} and then it is refined with {12, 30}. 5 CO2 K3
- (i) Create two clusters with each set of centroid mentioned above following the k-means approach.
- (ii) Calculate the SSE for each set of centroid.
- b. Explain in detail about Support Vector Machines with algorithms and state the role of hyper planes? 5 CO4 K3

(OR)

- c. Using the SVM algorithm, find the SVM classifier for the following data. 5 CO5 K3
- | Example No. | X_1 | X_2 | Class |
|-------------|-------|-------|-------|
| 1 | 3 | 1 | -1 |
| 2 | 4 | 2 | +1 |
- d. Given the following data, compute the principal component vectors and the first principal components: 5 CO4 K3
- | | x | 2 | 3 | 7 |
|---|---|---|----|---|
| y | 7 | 9 | 11 | |
- 5.a. Given the following data, compute the principal component vectors and the first principal components 5 CO2 K2
- | | x | -3 | 2 | -2 |
|---|---|----|---|----|
| y | 3 | -2 | 7 | |
- b. Given the dataset {a, b, c, d, e} and following distance matrix, construct a dendrogram by complete linkage hierarchical clustering using the agglomerative method 5 CO3 K3

	a	b	c	d	e	f
a	0	9	3	6	11	13
b	9	0	7	5	10	12
c	3	7	0	9	2	15
d	6	5	9	0	8	1
e	11	10	2	8	0	8
f	13	12	15	1	8	0

(OR)

- c. Explicate the weighted K-nearest Neighbour algorithm with a suitable sketch? 5 CO3 K2
- d. Explain KD Tree With an example in Machine Learning? 5 CO3 K3
- 6.a. Write the C4.5 Algorithm and Construct a Decision Tree Using C4.5 for the given dataset? 10 CO6 K2

Outlook	Humidity	Play
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes

(OR)

- c. Explain the AdaBoost algorithm in detail with its mathematical formulation, including the error calculation, weight updating mechanism, final hypothesis formulation, and its advantages and limitations? 5 CO5 K3
- d. Explain the Random Forest algorithm; including bagging and feature randomness, the tree construction process, Out-of-Bag (OOB) error, feature importance, and its advantages over decision trees? 5 CO6 K2

--- End of Paper ---