Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 27-10-2025 **Time:** 11.00am to 12.30pm

School: SOCSE	**Program:** M.Tech. in CSE specialization in Artificial Intelligence	
Course Code : AIE4002	**Course Name:** Machine Learning Algorithms	
Semester: I	**Max Marks**:50	**Weightage**:25%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**26**	**24**	**-**	**-**	**-**

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks. **5Q x 2M=10M**

1	Explain the purpose of splitting data into training and test sets.	**2 Marks**	**L2**	**CO1**
2	Illustrate the usage of regularization in machine learning models.	**2 Marks**	**L2**	**CO1**
3	Describe the cost function and its relation to model optimization.	**2 Marks**	**L2**	**CO1**
4	Summarize the usage of dendrogram in hierarchical clustering.	**2 Marks**	**L2**	**CO2**
5	Differentiate between agglomerative and divisive hierarchical clustering	**2 Marks**	**L2**	**CO2**

Part B

Answer the Questions. **Total Marks 40M**

6.	**a.**	Describe the concept of data scaling. Compare standard scaling and robust scaling techniques with examples, and explain why scaling is essential before training models.	**10 Marks**	**L2**	**CO 1**
		Or			
7.	**a.**	Discuss the structure and properties of a good dataset used for machine learning by stating the key characteristics that make a dataset suitable for training a model effectively?	**10 Marks**	**L2**	**CO 1**

8.	**a.**	Explain the data-generating process in machine learning. How does understanding the data source and its underlying distribution help in designing better models?	**10 Marks**	**L2**	**CO 1**
		Or			
9.	**a.**	Show how data is divided into training, validation, and test sets in machine learning. Discuss the importance of each set and elaborate on how cross-validation improves model reliability.	**10 Marks**	**L2**	**CO 1**

10.	**a.**	Describe how the Expectation-Maximization (EM) algorithm is used to estimate parameters in Gaussian Mixture Models (GMM).	**10 Marks**	**L3**	**CO 2**
		Or			
11.	**a.**	Illustrate with suitable diagrams how K-means++ improves the centroid initialization process and reduces convergence time by comparing between K-means and K-means++ algorithms.	**10 Marks**	**L3**	**CO 2**

12.	**a.**	Compare and contrast partition-based, hierarchical, and density-based clustering methods by highlight their advantages, disadvantages, and use cases with examples.	**10 Marks**	**L3**	**CO 2**
		Or			
13.	**a.**	How Spectral Clustering use graph Laplacians and eigenvalues to form clusters? Provide a suitable example.	**10 Marks**	**L3**	**CO 2**