Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Make Up Examinations – December 2025

Date: 26 – 12- 2025 **Time:** 1.00pm to 04.00pm

School: SOCSE	**Program:** B.TECH	
Course Code: CSE2028	**Course Name:** Statistical Foundation for data Science	
Semester: MK	**Max Marks**: 100	**Weightage**: 50%

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

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. **10Q x 2M=20M**

1.	Differentiate between supervised and unsupervised machine learning.	**2 Marks**	**L2**	**CO1**
2.	Explain the difference between a sample and a population in statistics.	**2 Marks**	**L1**	**CO1**
3.	Differentiate between L1 and L2 regularization in terms of their effect on model coefficients.	**2 Marks**	**L2**	**CO2**
4.	How does Lasso regularization help in feature selection?	**2 Marks**	**L1**	**CO2**
5.	Name any two clustering algorithms.	**2 Marks**	**L2**	**CO3**
6.	What is the main idea of hierarchical clustering?	**2 Marks**	**L1**	**CO3**
7.	Name two advantages of decision trees.	**2 Marks**	**L1**	**CO3**
8.	What is a neural network?	**2 Marks**	**L1**	**CO4**
9.	List any two layers commonly used in a CNN.	**2 Marks**	**L1**	**CO4**
10.	What does the GRU architecture solve compared to basic RNNs?	**2 Marks**	**L1**	**CO4**

Part B

Answer the Questions. **Total Marks 80M**

11.	**a.**	Classify and explain different types of data and variables with examples.	**10 Marks**	**L1**	**CO1**
	b.	Analyze the similarities and differences between sample surveys, experiments, and observational studies. Discuss their real-world applications with examples.	**10 Marks**	**L3**	**CO1**
		Or			
12.	**a.**	Explain the role and importance of descriptive statistics in data analysis.	**10 Marks**	**L2**	**CO1**
	b.	Explain how standard deviation and variance help in understanding data variability.	**10 Marks**	**L2**	**CO1**

13.	**a.**	Explain the steps involved in performing linear regression analysis.	**10 Marks**	**L1**	**CO2**
	b.	Explain how logistic regression is used in healthcare for outcome prediction with an example.	**10 Marks**	**L1**	**CO2**
		Or			
14.	**a.**	Discuss the steps in linear regression analysis, covering data preparation, model fitting, and performance evaluation.	**10 Marks**	**L2**	**CO2**
	b.	Explain how box plots and quartiles help identify outliers and skewness. Provide an example from sales data or test scores.	**10 Marks**	**L2**	**CO2**

15.	**a.**	Apply the concept of Support Vector Machines to a binary classification problem. Illustrate how the choice of kernel function influences the classifier's decision boundary.	**10 Marks**	**L3**	**CO3**
	b.	What is spectral clustering? Explain how it uses graphs and eigenvalues to perform clustering.	**10 Marks**	**L1**	**CO3**
		Or			
16.	**a.**	Apply the K-means clustering algorithm on a synthetic dataset with four clusters. Evaluate the effectiveness of the clustering using appropriate metrics such as inertia and silhouette score.	**20 Marks**	**L3**	**CO3**

17.	**a.**	Analyse how convolutional layers, pooling layers, and fully connected layers work together in a CNN to perform image classification.	**20 Marks**	**L3**	**CO4**
		Or			
18.	**a.**	What is deep learning? Explain its significance in solving complex real-world problems with examples.	**10 Marks**	**L3**	**CO4**
	b.	Write a short note on any two applications of deep learning such as sentiment analysis and recommender systems.	**10 Marks**	**L2**	**CO4**