Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 05/10/2026 **Time:** 02:00 PM - 03:30 PM

Course Code: CSD1712 **Course Name:** Statistical Foundations of Data Science

Semester: Third Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	60	120

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	What is Big Data?	2	CO1	BT2
2	What is dimensionality in data?	2	CO1	BT2
3	What is the impact of high dimensionality?	2	CO1	BT2
4	What is Multiple Linear Regression? Write its general equation.	2	CO2	BT2
5	State the Gauss-Markov Theorem.	2	CO2	BT2

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	Explain the role of Big Data and dimensionality in Economics and Finance.	10	CO1	BT2
7	Explain the rise of Big Data and dimensionality in Biological and Health Sciences.	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	Explain Big Data and discuss the major dimensions of Big Data.	10	CO1	BT2
9	The following data represents the number of hours studied X_1, hours of sleep X_2, and examination marks Y. **Student** | X_1 | X_2 | Y 1 | 2 | 6 | 55 2 | 3 | 7 | 60 3 | 4 | 6 | 65 4 | 5 | 8 | 72 5 | 6 | 7 | 75 Using the multiple linear regression model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ estimate the regression coefficients using the least-squares method and predict the marks for a student who studies 7 hours and sleeps for 8 hours.	10	CO2	BT3

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT

PTO (Please Turn Over)

10	The following response values are obtained: Y={4,9,16,25,36} Apply the Box-Cox transformation for: 1. $\lambda = 1$ 2. $\lambda = 0$ 3. $\lambda = 0.5$ Calculate the transformed values.	10	CO2	BT3
11	Given the observations: **X Y** 1 3 2 6 3 11 4 18 Assume a quadratic regression model: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ Determine the coefficients and predict Y when X=5	10	CO2	BT3

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	Consider a Kernel Ridge Regression model trained on two data points, with the following kernel matrix and target vector: K = [[1, 0.5], [0.5, 1]], y = [2, 4] The regularization parameter is λ = 0.5. (a) Write down the closed-form expression used to compute the coefficient vector α in kernel ridge regression, and briefly explain the role of the regularization term λI. (b) Compute the matrix (K + λI). (c) Find the inverse of (K + λI), showing your working (determinant and adjugate/cofactor method). (d) Hence calculate the coefficient vector α = $(K + \lambda I)^{-1}y$. (e) Suppose a new test point x* has kernel values k* = [k(x*, x_1), k(x*, x_2)] = [0.8, 0.3] with the training points. Using your computed α, find the predicted output $\hat{y}^* = k^{*T}\alpha$.	10	CO2	BT3
13	A Generalized Linear Model (Binomial Logistic Regression) relates the probability of success p to a predictor X as follows: log(p / (1−p)) = −2 + 0.8X Answer the following: (a) Explain briefly what "link function" is used here, and why it is needed in a GLM for binomial data. (b) Calculate the probability of success p when X = 2. (c) Calculate the probability of success p when X = 4. (d) Determine the odds ratio associated with a one-unit increase in X. (e) Interpret the results from (b), (c), and (d) in the context of how X affects the probability of success.	10	CO2	BT3