

CODE: 17CD05102

M. Tech I Year I Semester Regular Examinations, February 2018

FUNDAMENTALS OF DATA SCIENCE

(CSE)

Time: 3 hours

Max Marks: 60

Answer all **five** units. (5 x 12 = 60 Marks)

UNIT-I

1. (a) Briefly discuss the reasons for estimating the f .
(b) Write the commands in R Programming to compute the mean and variance of a vector of numbers and calculate standard deviation for the output

OR

2. (a) Write the commands in R Programming to select multiple rows and columns of a matrix at a time by providing data set as the indices.
(b) Explain two-step model-based parametric method approach to estimate f

UNIT-II

3. (a) Explain the process of using $\text{lm}()$ function to include interaction terms in a linear model.
(b) What is Residual Standard Error in assessing the quality of a linear regression fit model.

OR

4. (a) Briefly discuss the usage of $\text{lm}()$ function in accommodating non-linear transformations of the predictors.
(b) Explain the process of estimating the multiple linear regression Coefficients.

UNIT-III

5. (a) Explain the logistic model in Logistic Regression.
(b) Explain Linear Discriminant Analysis for the case $p=1$.

OR

6. (a) Explain the working principle of Quadratic Discriminant Analysis.
(b) State Bayes' Theorem used for classification.

Continued in page 2

UNIT-IV

7. (a) Give the comparison among full and Reduced QR Factorization
(b) Explain the functionalities of following functions in R programming subset(), apply(), sapply()

OR

8. (a) Explain Singular Value Decomposition by taking an example.
(b) Explain different stages in data wrangling

UNIT-V

9. (a) Briefly discuss Stars, Snowflakes, and Fact Constellations Schemas for Multidimensional Databases
(b) Explain different general optimization techniques for the efficient computation of data cubes

OR

10. (a) Give the comparison among Full Cube, Iceberg Cube Materialization
(b) What is data warehousing? Discuss the key features of data warehousing
