

Hall Ticket No

--	--	--	--	--	--	--	--	--	--

Question Paper Code: BCS001



INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)

M.Tech I Semester End Examinations (Supplementary) - January, 2019

Regulation: IARE-R16

FOUNDATIONS OF DATA SCIENCE

Time: 3 Hours

(CSE)

Max Marks: 70

Answer ONE Question from each Unit

All Questions Carry Equal Marks

All parts of the question must be answered in one place only

UNIT – I

1. (a) Draw the life cycle of data science project and summarize the stages of data science project. [7M]
(b) Describe how to explore data in R. Explain graphics and visualization in spotting problems. [7M]
2. (a) List the different R functions to read and write the data from disk and R object. Write R script to choose the character data dynamically from user. [7M]
(b) Explain and write the appropriate statements in R for the following operations. [7M]
 - Input as CSV File(input.csv)
 - Reading a CSV file
 - Apply different analysis on the file
 - Writing into a CSV File

UNIT – II

3. (a) Differentiate SQL and No SQL databases in detail. Give example of XML data extraction and operations using R. [7M]
(b) What is JSON file? Explain the input, output operations of JSON file using R. [7M]
4. (a) Compute covariance matrix and correlation matrix for the four numerical attributes. Interpret the statistical findings to know more about hidden nature in data. [7M]
(b) Summarize multiple regression in R and create equation for regression model. [7M]

UNIT – III

5. (a) How to predict whether an email is a spam and should be delivered to the junk folder. Suggest suitable data model. [7M]
- (b) Calculate the Jaccard coefficient for the given data [7M]
- $p = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0$
 $q = 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1.$
6. (a) Outline about the learning of a model? Write any four learning techniques and in each case give the expression for weight- updating. [7M]
- (b) Write short notes on Hierarchical clustering with `hclust()`. [7M]

UNIT – IV

7. (a) Give the basic structure of neural network and different artificial neural network with real time examples. [7M]
- (b) Discuss the difference of error in two hypotheses. Differentiate the MAP (maximum a posteriori) and ML (maximum likelihood) hypothesis. Give an example of a scenario in which a MAP hypothesis is preferable to an ML hypothesis. [7M]
8. (a) Describe the prediction model in terms of the following measures for best fit: Residual standard error, Multiple R-squared, F-statistic, p-value [7M]
- (b) Compare the learning algorithms with example in terms of problem nature, accuracy and error rate. [7M]

UNIT – V

9. (a) Write R script to plot a data frame having: `{df1:{red,green,blue,pink,black} df2: {3,5,8,10,34}}` using relevant plot. [7M]
- (b) Generalize the graphical analysis in data analysis. List the various plots in R and explain in detail. [7M]
10. (a) How would you get the multiple plots in single window? Plot the regression model along with residuals. Write a R script for creating a boxplot of iris sepal length attribute. [7M]
- (b) Elaborate how to export a graph using graphics parameters. How to export the text data to plot with example. [7M]