



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING

TUTORIAL QUESTION BANK

Course Name	Foundations of Data Science
Course Code	BCS001
Class	I M. Tech
Branch	Computer Science Engineering
Year	2018 - 2019
Team of Instructors	Mr. C Raghavendra

OBJECTIVES

To meet the challenge of ensuring excellence in engineering education, the issue of quality needs to be addressed, debated and taken forward in a systematic manner. Accreditation is the principal means of quality assurance in higher education. The major emphasis of accreditation process is to measure the outcomes of the program that is being accredited.

In line with this, faculty of Institute of Aeronautical Engineering, Hyderabad has taken a lead in incorporating philosophy of outcome based education in the process of problem solving and career development. So, all students of the institute should understand the depth and approach of course to be taught through this question bank, which will enhance learner's learning process.

PART – A (SHORT ANSWER QUESTIONS)

S. No	QUESTIONS	Blooms taxonomy level	Course outcome
UNIT – I INTRODUCTION			
1.	List the major steps involved in data science project.	Remember	1
2.	Identify the steps in creating and running a R script?	Remember	2
3.	Write the steps to create a calculator application with R objects?	Understand	3
4.	Write the basic set of steps followed in data management?	Understand	2
5.	State the any few data sources in data collection step.	Remember	2
6.	Draw the block diagram of data science project process.	Understand	2
7.	State the features of R Language.		
8.	Write the various classes of data types ANSI C supports?	Remember	4
9.	State which of the following are valid identifiers. If not invalid, state the reason. a. q b. C c. c d. t e. pi	Understand	3

S. No	QUESTIONS	Blooms taxonomy level	Course outcome
10.	Write expression in R a) $e^{4+\log 2}$ b) 24×53 c) $\log_2 10$ d) $\log_{10} 2$	Remember	3
11.	What are the different forms of data types and how to test the data type in R? Give one example for each	Understand	3
12.	Diffentiate R while comparing with other programming languages?	Remember	3
13.	How to set up the R environment?	Understand	4
14.	Explain R as calculator using basic operations and inbuilt functions with suitable example?	Remember	3
15.	What are different basic components in R?	Remember	3

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – I INTRODUCTION			
1.	Elaborate the following R objects. a)vector b)data frame c)matrix d)list	Understand	2
2.	Compute the given <u>mathematical</u> formula and display on console in R. $X=$	Remember	3
3.	Describe the multiple ways to read and write data from disc and web. Elaborate R functions to choose character, numerical input dynamically from user.	Understand	1
4.	List the inbuilt summary functions to apply on vectors. Create a vector and apply all functions on it.	Remember	3
5.	Identify the different ways to access the R objects. List the different data types in R with suitable example.	Remember	1
6.	The price of one kg of rice is Rs. 40.75 and one kg of sugar is Rs. 30. Write R program to get the total amount of 2kg rice and 5kg sugar purchase.	Understand	4
7.	Elaborate the process of working with different data files. Write about filechoose() function.	Remember	2
8.	List the inbuilt summary functions to apply on vectors. Create a vector and apply all functions on it.	Understand	2
9.	How to get system date in R? Generate sequence of previous and coming 10 dates from today in R.	Understand	3
10.	What are the different ways to read the dataset? How to create and rename a variable in R? What are the read write methods available in R and explain?	Understand	3

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome																								
UNIT – I INTRODUCTION																											
1.	Create an Array with name “MySales” with 30 observations using following methods: a) By using the array with dimensions 3, 5 and 2. b) By using Vector method.	Remember	2																								
2.	What is data frame and how to create a data frame using the following data: Height GPA 66 3.80 62 3.78 63 3.88 70 3.72 74 3.69	Remember	4																								
3.	The ABC electric company manufactures four consumer products. Their inventory position on a particular day is given below. Code Quantity Rate(Rs.) F105 275 575.00 H220 107 993.95 I019 321 215.50 M315 89 725.00 Write a R program to prepare the inventory report table in the following format: INVENTORY REPORT <table style="width: 100%; border-collapse: collapse;"><thead><tr><th style="text-align: left;">Code</th><th style="text-align: left;">Quantity</th><th style="text-align: left;">Rate</th><th style="text-align: left;">Value</th></tr></thead><tbody><tr><td>---</td><td>---</td><td>---</td><td>---</td></tr><tr><td>---</td><td>---</td><td>---</td><td>---</td></tr><tr><td>---</td><td>---</td><td>---</td><td>---</td></tr><tr><td>---</td><td>---</td><td>---</td><td>---</td></tr><tr><td>Total Value:</td><td>-----</td><td></td><td></td></tr></tbody></table>	Code	Quantity	Rate	Value	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Total Value:	-----			Remember	3
Code	Quantity	Rate	Value																								
---	---	---	---																								
---	---	---	---																								
---	---	---	---																								
---	---	---	---																								
Total Value:	-----																										
4.	How will you identify and treat the missing value and outlier data in R?	Remember	5																								
5.	Compare the different forms of data types and create a list object with restaurant menu items. Give inbuilt functions on list object with example Write R script to create and display list object of stores items having: {Fruits:{orange,mango,apple,watermelon,banana}}Juices:{appy,fruty, slice} Milkshakes:{Mango, papaya, sapota, pineapple}}.	Remember	4																								
6.	Write the R script for the following: i) Import the csv file studentmarks.txt having sno,sem_no,m1.m2,m3,m4,m5,m6,L1,L2 using R ii) Export the Data Frame Smarks to CSV files. iii) Create a new variable Totalmarks and avgmarks with sum and avg of m1, m2, m3, m4, m5, m6. iv) Check for Outliers in the studentmarks data data frame. v) Create a new data set by combining 2 data sets. vi)Check for missing data in studentmarks and remove missing data.	Remember	5																								

7.	Let two vectors $x \leftarrow c(1,3, 5)$ and $y \leftarrow c(3, 2, 10)$, what is size of the expression <code>rbind(x, y)</code> output? Differences between <code>rbind</code> and <code>cbind</code> . Create a matrix using <code>rbind</code> and <code>cbind</code> .	Remember	5
8.	Create a data frame with $a = c(1, 2, 3)$, $b = c(4, 5, 6)$, $c(7, 8, 9)$ and find the value of the following 1a.) How do I select the $c(4, 5, 6)$? 1b.) How do I select the 1? 1c.) How do I select the 5? 1d.) What is <code>df[, 3]</code> ? 1e.) What is <code>df[1,]</code> ? 1f.) What is <code>df[2, 2]</code> ?	Understand	8
9.	Find the output in R: i) If x is defined as $x \leftarrow \text{list}(2, "a", "b", \text{TRUE})$. What does <code>x[[1]]</code> ? ii) If x is defined as $x \leftarrow \text{list}(2, "a", "b", \text{TRUE})$. What does <code>x[[2]]</code> give me? iii) if x, y are two vectors $x \leftarrow 1:4$ and $y \leftarrow 2:3$. What is produced by the expression <code>x + y</code> ? iv) if x, y are two vectors, $x \leftarrow 1:4$ and a vector $y \leftarrow 2$. What is produced by the expression <code>x + y</code> ? v) Find x^3	Understand	9
10.	Analyze the control structures with conditional statements in R with suitable example. Create a user defined function <code>fact(j)</code> to return the the factorial of j using functions in R.	Understand	8

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – II SQL, NOSQL AND DATA ANALYSIS			
1.	State the packages and function required to run SQL queries in R?	Understand	3
2.	List the packages and function required to read Excel sheet data to R object?	Understand	5
3.	Explain the drawbacks in different ways to access Excel sheet in R.	Understand	3
4.	What are other names of NoSQL?	Understand	6
5.	List the features of NoSQL.	Understand	3
6.	Give any five examples of NoSQL databases.	Understand	7
7.	Differentiate SQL and No SQL databases.	Understand	4
8.	Compare the SQL and NoSQL in terms of Data storage model and Schema.	Remember	5
9.	What are the assumptions of regression Modelling	Remember	7
10.	What is ANOVA?	Understand	3

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – II SQL, NOSQL AND DATA ANALYSIS			
1.	Write the R script which include relevant packages and procedure to access .csv and .exl files. Elaborate with suitable example.	Understand	4
2.	What is the difference between SQL database and NoSQL Database?	Remember	3
3.	Explain different approaches in R to connect with Excel	Understand	5

4.	Define basic classification based on data model, with examples.	Remember	6
5.	Explain Covariance and Correlation with example. Infer the relation among attributes with respect to correlation coefficient.	Remember	4
6.	Calculate the correlation coefficient of iris dataset. What preprocess steps is needed for iris data set.	Understand	5
7.	Write a R program to find the correlation coefficient of iris data set and find the exact relation by using linear regression model.	Understand	5
8.	Distinguish simple and multiple regression analysis and its applications working with numerical and categorical data?	Remember	7
9.	Differentiate SQL and No SQL databases in detail?	Understand	6
10.	What is forecasting give examples?	Remember	5

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – II SQL, NOSQL AND DATA ANALYSIS			
1.	Write R program to extract sampleXML data from web and steps to convert as dataframe. Specify the needed packages and functions.	Remember	9
2.	Distinguish simple and multiple regression analysis and its applications working with numerical and categorical data?	Remember	8
3.	What are residuals? Define in Regression analysis.	Understand	9
4.	Generate prediction model using linear regression for finding relative relation among variables. Write a R script to get a linear equation $y=mx+c$ form for the heart weight and body weight in cats dataset.	Remember	8
5.	Compute the covariance matrix and correlation matrix for the four numerical attributes. Interpret the statistical findings to know more about hidden nature in data.	Remember	9
6.	Write an R script to connect with Excel, read the contents of sheet and load into R object.	Remember	8
7.	Write a R program to perform the following: i. Import a data from web storage. ii. Name the dataset with suitable identifier iii. Perform Logistic Regression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. iv. Check the model is fit or not.	Remember	9
8.	Write R program to perform the following: i. Find the correlation matrix of iris data set ii. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data. iii. Perform analysis of covariance	Remember	9

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – III DATA MODELS			
1.	State the applications of regression models?	Remember	3, 4
2.	List various types of regression models?	Understand	4
3.	State the difference between numerical and categorical parameters?	Understand	3
4.	Justify a single model on data is suggestible.	Remember	3
5.	State the various types of functions in R to support linear regression?	Remember	3
6.	State the various attributes to evaluate the multiple regression?	Remember	4
7.	State the residuals impact in linear model.	Remember	4
8.	What are the evaluating measures in regression models.	Remember	4
9.	What is the role of machine learning algorithms in data model.	Remember	3
10.	State the steps to evaluate the data model.	Understand	3

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – III DATA MODELS			
1.	Outline about the learning of a model? Write any four learning techniques and in each case give the expression for weight- updating.	Remember	3
2.	Describe the limitations of the perception model. How to create and evaluate a data model. Describe with one case study.	Understand	3
3.	List out the applications of Machine learning with example. Discuss about a consistent learner and what it means for a set of training examples to be linearly separable.	Understand	3
4.	Write the steps to describe the process to create and evaluate the data model for the given data.	Remember	4
5.	Predict whether an email is a spam and should be delivered to the Junk folder. Suggest suitable data model.	Remember	4
6.	Discuss about clustering in detail? Elaborate k means algorithm.	Understand	7
7.	Write a R script to implement	Remember	6
8.	State about conditional probability in the bayes theorem for different type of data classification problem with suitable example.	Remember	8
9.	Calculate the Jaccard coefficient for the given data $p = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0$, $q = 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1$;	Remember	6
10.	List the different types clustering. Write about k-nn algorithm. Write a R script to cluster the mtcars dataset using k-nn algorithm.	Understand	5

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – III DATA MODELS			
1.	Find the suitable Machine learning algorithm for the following real time problems: i. Predicting Housing Prices ii. Analyzing Sentiment iii. Finding Similar Documents iv. Recommending Products	Understand	3
2.	How can you identify, the best fit data model from the given dataset.	Remember	4
3.	Describe the prediction model in terms of the following measures for best fit: Residual standard error, Multiple R-squared, F-statistic, p-value	Remember	4
4.	State the suitable classification algorithm for pima Indian diabetes dataset classification. Improve the accuracy by performing multiple preprocess steps.	Understand	7

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – IV ARTIFICIAL NEURAL NETWORKS			
1.	Draw the basic structure of neural network.	Remember	3, 4
2.	List various types of problems for neural network learning?	Understand	4
3.	State the difference between neural network and multi-layer neural network?	Understand	3
4.	What is meant by perceptions in neural network?	Remember	3
5.	State the various types of layers in neural network design?	Remember	3
6.	State the remarks in multi-layer perceptions?	Remember	4
7.	State the role of error function in neural network.	Remember	4
8.	What are the evaluating measures of neural network model?	Remember	4
9.	Define Multi-layer networks	Remember	3
10.	Give a short note on feed forward networks	Understand	3

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – IV ARTIFICIAL NEURAL NETWORKS			
1.	Outline about the learning of a model? Write any four learning techniques and in each case give the expression for weight- updating.	Remember	9
2.	Describe the limitations of the perception model. How to create and evaluate a data model. Describe with one case study.	Understand	8
3.	List out the applications of Machine learning with example. Discuss about a consistent learner and what it means for a set of training examples to be linearly separable.	Understand	10
4.	Write the steps to describe the process to create and evaluate the data model for the given data.	Remember	11
5.	Predict whether an email is a spam and should be delivered to the Junk folder. Suggest suitable data model.	Remember	8

6.	Give the basic structure of neural network and different types of ANN with real time examples.	Remember	7
7.	State the suitable classification algorithm for pima Indian diabetes dataset classification. Improve the accuracy by performing multiple preprocess steps.	Understand	9
8.	How to evaluate hypothesis of different types with examples? Explain the basics of sampling theory.	Understand	10
9.	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.	Remember	10
10.	Compare the learning algorithms with example in terms of problem nature, accuracy and error rate.	Remember	11

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – IV ARTIFICIAL NEURAL NETWORKS			
1.	Find the suitable Machine learning algorithm for the following real time problems: i. Predicting Housing Prices ii. Analyzing Sentiment iii. Finding Similar Documents iv. Recommending Products	Understand	13
2.	How can you identify, the best fit data model from the given dataset.	Understand	11
3.	Describe the prediction model in terms of the following measures for best fit: Residual standard error, Multiple R-squared, F-statistic, p-value	Understand	9
4.	How to evaluate hypothesis of the given problem. Describe the basic principle of sampling theory.	Understand	12
5.	List the steps to improve the accuracy of neural network data model.	Remember	11

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – V DELIVERING RESULTS			
1.	State the applications of regression models?	Remember	3, 4
2.	List various types of regression models?	Understand	4
3.	State the difference between numerical and categorical parameters?	Understand	3
4.	Justify a single model on data is suggestible.	Remember	3
5.	State the various types of functions in R to support linear regression?	Remember	3
6.	State the various attributes to evaluate the multiple regression?	Remember	4
7.	State the residuals impact in linear model.	Remember	4
8.	What are the evaluating measures in regression models.	Remember	4
9.	What is the role of machine learning algorithms in data model.	Remember	3

10.	State the steps to evaluate the data model.	Understand	3
-----	--	------------	---

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – V DELIVERING RESULTS			
1.	Generalize the graphical analysis in data analysis? List the various plots in R and explain in detail.	Remember	14
2.	How to plot the word (text) data based on frequency of words. Write R script to plot a data frame having: {df1: {sea,river,pond,lake,pool} df2: {3,5,8,10,34}} using relevant plot.	Understand	12
3.	List out the applications of Machine learning with example. Discuss about a consistent learner and what it means for a set of training examples to be linearly separable.	Understand	9
4.	Write the steps to plot the numerical data at different granularity levels.	Remember	10
5.	Plot the k-means clustering algorithm outcome. State the relevant packages and functions in R	Remember	7
6.	How would you get the multiple plots in single window?	Remember	13
7.	Elaborate how to export a graph using graphics parameters. How to export the text data to plot with example.	Understand	11
8.	Describe boxplot() of iris\$petal length attribute. Specify the observations of plot.	Understand	12
9.	Plot the regression model along with residuals.	Remember	15
10.	Write a R script for creating a boxplot of iris sepal length attribute.	Remember	14

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
UNIT – V DELIVERING RESULTS			
1.	Describe the various plots in R to visualize the data and explain the purpose of each plot in detail.	Understand	13
2.	Write R script to plot a data frame having: {df1: {red,green,blue,pink,black} df2: {3,5,8,10,34}} using relevant plot.	Understand	11
3.	List out the steps to plot the data models with relevant packages.	Remember	14
4.	How to partition the window to get more number of plots. Discuss on single and multi object plots in R.	Understand	13
5.	Discuss about the residuals with respect to observed values? State a case study to show the fitted line and residuals in logistic regression.	Understand	9

Prepared by: Mr. C Raghavendra, Assistant Professor.

HOD, CSE