Time: 2 hours   Maximum marks: 20

Answer all questions.

1. How does the level of lateral acceleration, which represents the sideways forces experienced during bus turns primarily controlled by the driver, impact the level of discomfort or motion sickness as perceived by bus passengers? A study reported data on x, denoting the degree of lateral acceleration (measured in accordance with a British standard for assessing similar motion at sea), and y, indicating the self-reported discomfort level (%).

   Values of the degree of lateral acceleration in the sample ranged from 6.0 to 17.6.

   Relevant summary statistics are also available.

   $n = 17$, $\Sigma x_i = 222.1$, $\Sigma y_i = 193$, $\Sigma x_i^2 = 3056.69$, $\Sigma y_i^2 = 2975$, $\Sigma x_i y_i = 2759.6$.

   (a) Obtain the simple linear regression model relating these two variables (supported by the raw data).   (3 marks)

   (b) Interpret the estimate of the slope parameter that conveys information about the precision and reliability of estimation.   (1 mark)

   (c) Would it be sensible to use the simple linear regression model as a basis for predicting the discomfort level (%) when the the degree of lateral acceleration $= 5.0$? Explain your reasoning.   (1 mark)

2. Within the domain of oil drilling operations, various components within the drilling assembly face the potential of sulfide stress cracking. An analysis of the chemical composition of a standard steel grade was carried out, yielding data concerning two critical parameters: y, the threshold stress (% SMYS), and x, the yield strength (MPa).

x	635	644	711	708	836	820	810	870	856	923	878	937	948
y	100	93	88	84	77	75	74	63	57	55	47	43	38

   Relevant summary quantities are

   $\Sigma x_i = 10,576$, $\Sigma y_i = 894$, $\Sigma x_i^2 = 8,741,264$, $\Sigma y_i^2 = 66,224$, $\Sigma x_i y_i = 703,192$.

   (a) Determine and interpret the coefficient of determination.   (4 marks)

(b) Calculate a confidence interval using confidence level 95% for the expected change in stress associated with a 1 MPa increase in strength. (4 marks)

(c) Carry out the model utility test at level .05, using the rejection region approach. (2 marks)

3. The primary durability concern in reinforced concrete structures is the corrosion of steel reinforcing bars. Concrete carbonation, a chemical process, leads to a reduction in the pH value, which can trigger the corrosion of the rebar. The provided dataset consists of x, representing the depth of carbonation in millimeters, and y, representing the strength in megapascals (MPa) of a selection of core specimens extracted from a specific building.

x	8	15	16.5	20	20	27.5	30	30	35
y	22.8	27.2	23.7	17.1	21.5	18.6	16.1	23.4	13.4
x	38	40	45	50	50	55	55	59	65
y	19.5	12.4	13.2	11.4	10.3	14.1	9.7	12	6.8

Relevant quantities are as follows:

$\Sigma x_i = 659.0$, $\Sigma x_i^2 = 28,967.50$, $\bar{x} = 36.6111$, $\Sigma y_i = 293.2$, $\Sigma x_i y_i = 9293.95$.
$\Sigma y_i^2 = 5335.76$, $\hat{\beta}_1 = -0.297561$, $\hat{\beta}_0 = 27.182936$, $SSE = 131.2402$,
$r^2 = 0.766$, $\hat{\sigma}$ or $s = 2.8640$.

(a) Compute a 95% confidence interval to estimate the true average strength of all core specimens with a carbonation depth of 45mm. (2 marks)

(b) Calculate a 95% prediction interval for a strength value that would result from selecting a single core specimen whose depth is 45mm. (2 marks)

(c) Compare the two intervals obtained in parts (a) and (b). Explain the difference between them. (1 mark)

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

Programme: M.Sc. (Five Year Integrated) in Computer Science (Artificial Intelligence & Data Science) Course Code & Title: 21-805-0502 Big Data Analytics			
Name of Examination: Series I		Max. Marks: 20	Semester: V
Batch: 2021-26	Duration: 2 Hours	Date: 30.10.2023	Time: 10:00 AM – 12 PM

Answer all questions

1. Explain KNN algorithm for classification. Also explain how it can be used for regression
   (3 marks)

2. Create a decision tree for the following data   (4 marks)

A	B	C	Target
a	X	u	P
a	X	v	P
a	Y	u	P
b	Y	v	N
b	Y	u	N
b	X	v	P

3. Explain different activation function used in Neural network   (3 marks)
4. Explain different characteristics of Big data   (2 marks)
5. Give examples for structured, unstructured and semi structured data   (2 marks)
6. Suppose you have a large text file that contains the names and ratings of different movies. Each line in the file has the following format: `movie_name, rating` where `movie_name` is a string and rating is an integer between 1 and 5. You want to use MapReduce to find the average rating for each movie. Draw a pictorial representation for the map and reduce functions that can accomplish this task. Explain the logic behind your picture and how it works.   (4 marks)
7. Draw and explain the Hadoop cluster architecture   (2 marks)

Department of Computer Science, Cochin University of Science & Technology
M. Sc. Computer Science (Artificial Intelligence and Data Science)
First Internal Examination, October 2023. Total: 20 Marks

21-805-0503: CLOUD COMPUTING

Section I: Choose the best answer among the options given (6 marks)

1. A webs streaming platform is an example of which computing model?
a) IaaS   b) PaaS   c) SaaS   d) DaaS

2. Which among the following is NOT an advantage of moving to private cloud?
a) Pay as you Use model   b) Vendor lock in   c) Scalability   d) Reliability

3. Which among the following is a deterrent to attacks from automated bots on websites?
a) Honeypot captcha   b) Strength of Passwords   c) lock and key   d) Rate limiting

4. Which among the following is NOT a security objective of cloud computing?
a) Confidentiality   b) Integrity   c) Accessibility   d) Controllability

5. Which among the following models provides SDKs for building several applications in a closed environment?
a) IaaS   b) PaaS   c) SaaS   d) DaaS

6. NoSQL is suitable for which among the following scenarios?
a) store small amount of data in structured format
b) store large amount of data in unstructured format
c) store small amount of data in unstructured format
d) store large amount of data in structured format

Section II: Answer the following (Total 14 Marks)

7. What are the disadvantages of cloud computing?   (3 marks)

8. What is a multi-tenant SaaS application?   (3 marks)

9. Describe a Man in the Middle attack and provide solutions against it   (4 marks)

10. How does streaming data gets shared using MQTT protocol?   (4 marks)

****

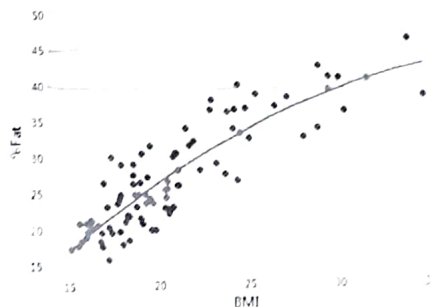|	COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY DEPARTMENT OF COMPUTER SCIENCE		
Programme: M.Sc. (Five Year Integrated) in Computer Science (Artificial Intelligence & Data Science) **Course Code & Title: 21-805-0504 R for Data Science**
	Name of Examination: Series I	Max. Marks: 20	Semester: V
Batch: 2021-26	Duration: 2 Hours	Date: 31.10.2023	Time: 10:00 AM – 12 PM

Answer all questions

1. Consider the two expressions given below. Are they same? Do they return the same answer
   foo<-list(1,2,3,4)
   faa<- list(1:4)   (2 marks)

2. Write a program in R that takes a vector of numbers as input and returns the sum of the squares of the numbers. Demonstrate the working of the code with the vector c(1, 2, 3, 4, 5).
                                                                                   (2 marks)

3. Write a *for* loop in R that prints the numbers from 1 to 10 in reverse order. (2 marks)

4. Write an ifelse statement in R that assigns the value "positive" to a variable x if a given number y is greater than zero, "negative" if y is less than zero, and "zero" if y is equal to zero.
                                                                                   (3 marks)

5. Write a code in R that creates a matrix m with 3 rows and 4 columns, filled with random numbers between 0 and 1. Then, write a code that extracts the second column of m as a vector v.
                                                                                   (3 marks)

6. Write a code in R that converts a character vector c("red", "green", "blue", "yellow") into a factor variable f with levels ordered alphabetically.   (3 marks)

7. Write R script for a bar plot of the number of cylinders (cyl) in the mtcars dataset. Add appropriate title, labels, and legend to the plot.   (3 marks)

8. Give any 3 data visualisation tools including an example script. (3 marks)

9. During an experiment, I measured the Body Mass Index (BMI) and body fat percentage of adolescent girls. I graphed these two variables in a scatterplot to assess the relationship between them. What can you infer from the plot.   (3 marks)

Time: 2 hours   Maximum marks: 20

Answer all questions.

1. Explain the method of least squares and how it is used to estimate the regression coefficients in multiple linear regression. What is the interpretation of the regression coefficients in the context of multiple linear regression?   (4 marks)

2. Explain the principles behind Principal Component Regression (PCR). (3 marks)

3. Discuss the concept of subset selection of independent variables and its impact on model performance.   (3 marks)

4. A bottler of soft drinks is examining the service routes of vending machines within their distribution system. The primary goal is to predict the time it takes for route drivers to service vending machines at different outlets. This service task involves restocking the machines with beverage products and performing minor maintenance or housekeeping. The industrial engineer overseeing the study proposes that the two most crucial factors influencing delivery time (y) are the number of cases of products stocked (x_1) and the distance walked by the route driver (x_2). The engineer has gathered 25 observations on delivery time. Let X and y be the $25*3$ matrix of the levels of the regressor variables and $25*1$ matrix of the observations, respectively. The least-squares fit (with the regression coefficients reported to five decimals) is

$$\hat{y} = 2.34123 + 1.61591 x_1 + 0.01438 x_2.$$

Required informations for computations:

$$(X'X)^{-1} = \begin{bmatrix} 25 & 219 & 10232 \\ 219 & 3055 & 133899 \\ 10232 & 133899 & 6725688 \end{bmatrix}^{-1} = \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 559.60 \\ 7375.44 \\ 337072.0 \end{bmatrix} , \hat{\sigma}^2 = 106239$$

Now, using the provided matrices and regression equation, answer the following questions:

(a) Find a 95% confidence interval for the parameter β_1. (3 marks)

(b) Construct a 95% confidence interval on the mean delivery time for an outlet requiring $x_1 = 8$ cases and where the distance $x_2 = 275$ feet. (4 marks)

(c) Construct a 95% prediction interval on the delivery time at an outlet requiring $x_1 = 8$ cases and where the distance $x_2 = 275$ feet. (3 marks)

M.Sc. (Five-Year Integrated) in Computer Science (Artificial Intelligence & Data Science)

Series II Examination - November 2023

21-805-0502: Big Data Analytics

Duration : 2 Hrs   Maximum Marks : 20

1. What are the challenges and benefits of Big Data Analytics

[3 marks][CO1. DL-1. BTL-2]

2. What are the advantages of choosing HDFS for Big Data Analytics

[2 marks][CO5. DL-1. BTL-2]

3. How does Spark handle data storage and processing?

[3 marks][CO5. DL-1. BTL-2]

4. What are Resilient Distributed Datasets (RDDs) in spark?   [3 marks][CO5. DL-1, BTL-2]

5. How structured streaming is handled in spark [5 marks][CO5, DL-2, BTL-3]

6. Explain Support Vector Machine. How it supports non linearity

[4 marks][CO4. DL-2. BTL-3]

21-805-0503: CLOUD COMPUTING

Answer the following (5 Marks each)

1. What is a hypervisor? What are its different types?   [CO6, DL-1, BTL-1]

2. What is the difference between full virtualization and para-virtualization?
                                                                  [CO7, DL-1, BTL-2]

3. How are containers reducing capital expenditure in computing?   [CO7, DL-2, BTL-2]

4. Write an API to accept user credentials and return authentication result. [CO5, DL-3, BTL-3]

****

Cochin University of Science and Technology
Department of Computer Science

M.Sc. (Five-Year Integrated) in Computer Science (Artificial Intelligence & Data Science)

Series II Examination - November 2023

21-805-0204: R for Data Science

Duration : 2 Hrs   Maximum Marks : 20

---

1. What are the features of Time Series data. How Time Series Analysis is implemented in R
   [3 marks][CO1, DL-1, BTL-2]

2. How multiple linear regression is implemented in R   [4 marks][CO1, DL-1, BTL-2]

3. How Object oriented programming is implemented in R?
   [3 marks][CO, DL-1, BTL-2]

4. Using the following data, perform a oneway analysis of variance using $=.05$. .

Group1	Group2	Group3
51	23	56
45	43	76
33	23	74
45	43	87
67	45	56

(1)

F(2,12)=3.89 from the table   [10 marks][CO3, DL-3, BTL-4]

1

Time: 2 hours   Total marks: 20

Answer all questions.

1. Find the continued fraction of $\sqrt{5}$ and $\sqrt{3}$ (2 marks)

2. In a RSA cryptosystem a particular A uses two prime numbers P=13 and Q=17, to generate his public and private keys. If the public key of A is 35, then the private key of A is_____.   (2 marks)

3. Show that 1729 is a pseudoprime?   (2 marks)

4. Explain the knapsack algorithm with example.   (4 marks)

5. Explain the Diffie-Hellman Key Exchange Algorithm with example.   (3 marks)

6. Explain the El Gamal Asymmetric Key cryptography method.   (3 marks)

7. Find a factor of n=1189 using pollard 's rho method , set X_0=2 and f(x)=$x^2 + 1$

                                                        (4 marks)