

Tribhuvan University  
Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
First Assessment 2080

Subject: Mathematics for Data Science  
Course No: MDS 504  
Level: MDS /I Year /I Semester

Full Marks: 45  
Pass Marks: 22.5  
Time: 2 hrs

Candidates are required to give their answer in their own words as far as practicable.

Attempt All Questions:

**Group A [5×3 = 15]**

1. What is the parallel coordinates method? Explain with an example. What is the use of this method in data science?
2. Define  $L_1$ ,  $L_2$  and  $L_\infty$  norms on  $\mathbb{R}^n$ . Calculate  $L_1$ ,  $L_2$  and  $L_\infty$  norms of the vector  $x = (1, -1, 0, \dots, 0, 2)$  on  $\mathbb{R}^n$ .
3. What is the angle between the diagonal of the unit cube in the positive orthant and the vector  $e_1$  in  $\mathbb{R}^3$ ?
4. Show that  $u = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$  and  $v = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$  are orthogonal in  $\mathbb{R}^2$  and find corresponding orthonormal basis for  $\mathbb{R}^2$ .
5. Prove that eigen vectors  $v_1$  and  $v_2$  that correspond to distinct eigen values  $\lambda_1$  and  $\lambda_2$  of a  $2 \times 2$  matrix are linearly independent.

**Group B [5×6 = 30]**

6. Let  $S$  and  $T$  be matrix transformations defined by  $S(y) = Ay$  and  $T(x) = Bx$ , where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 0 \\ 5 & -2 \\ 0 & 1 \end{pmatrix}.$$

- a) What are the domains and codomains of  $S$  and  $T$ ? Why is the composite transformation  $S \circ T$  defined? What are the domain and the codomain of  $S \circ T$ ?
- b) Let  $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ . Determine  $T(x)$ .
- c) Find  $(S \circ T)(x)$ .
- d) Find a matrix  $C$  so that  $(S \circ T)(x) = Cx$ .
- e) Show that  $S \circ T$  is linear.

7. Prove that if  $\theta$  is the angle between two non-zero vectors  $x, y$  in  $\mathbb{R}^n$ , then

$$x \cdot y = \|x\| \|y\| \cos \theta.$$

8. Let

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}, \quad B = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

Show that  $V$  is a subspace of  $\mathbb{R}^3$ , and  $B$  is a basis for  $V$ .

OR

Show that the following set of vectors is a basis for  $\mathbb{R}^3$ , and then express the standard basis vectors:  $e_1, e_2, e_3$  in terms of these:

$$u_1 = \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}, u_2 = \begin{pmatrix} -5 \\ 2 \\ 3 \end{pmatrix}, u_3 = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}.$$

9. Consider the following matrix:  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ .

- a) What can we say about the action of  $A$  on an arbitrary vector?
- b) What are examples of eigen values and eigen vectors of this matrix?
- c) What does the discussion for this example illustrate?

OR

Find the eigenvalues and eigenvectors of  $A = \begin{pmatrix} 5 & -2 \\ 7 & -4 \end{pmatrix}$ . State and sketch the effect of multiplying the eigen vector by the matrix  $A$ .

10. Let  $A$  be a square matrix with eigen vector  $u$  belonging to Eigen value  $\lambda$ . Prove that

- a) If  $m$  is a natural number then  $\lambda^m$  is an eigen value of the matrix  $A^m$  with the *same* eigen vector  $u$ .
- b) If the matrix  $A$  is invertible then the eigen value of the inverse matrix  $A^{-1}$  is  $1/\lambda = \lambda^{-1}$  with the *same* eigen vector  $u$ .



Tribhuvan University  
Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
First Assessment 2080

**Subject: Data Structure and Algorithms**  
**Course No: MDS 502**  
**Level: MDS /I Year /I Semester**

**Full Marks: 45**  
**Pass Marks: 22.5**  
**Time: 2 hrs**

*Candidates are required to give their answer in their own words as far as practicable.*  
**Attempt All Questions.**

**Group A [5×3 = 15]**

1. Compare ADT with data type. Explain big oh (O) notation with example. (1.5 + 1.5)
2. Explain push and pop operations of stack. (1.5 + 1.5)
3. What are different applications of queue?
4. Compare recursion with iteration. Write a recursive function to find  $n^{\text{th}}$  Fibonacci number. (1.5 + 1.5)
5. Compare linked list with array. What is list ADT? (2 + 1)

**Group B [5×6 = 30]**

6. Explain algorithm to convert an infix expression to postfix. Use this algorithm to convert the infix expression  $(A + B) * (C - D) / E$  to postfix. (3 + 3)

**OR**

How do you implement stack using linked list? Explain. (6)

7. Explain priority queue in detail. What are different ways to implement priority queue? (2 + 4)

**OR**

Define queue. How do you implement queue operations using in array structure? Explain. (1 + 5)

8. Define tail recursion. Explain recursive algorithm to solve Tower of Hanoi problem. (2 + 4)

9. How can you insert and-remove elements in singly linked list? Explain. (6)

10. What is header node in linked list? Explain circular linked list with example. (2 + 4)

\*\*\*

AB + CD - E / F

Tribhuvan University  
Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
First Assessment 2080

**Subject: Fundamentals of Data Science**  
**Course No: MDS 501**  
**Level: MDS /I Year/I Semester**

**Full Marks: 45**  
**Pass Marks: 22.5**  
**Time: 2 hrs**

*Candidates are required to give their answers in their own words as far as practicable.*  
**Attempt ALL questions.**

**Group A**                      **[5 × 3 = 15]**

1. Why is Data Science often regarded as a role with blurry or ambiguous boundaries? Provide rationale to support your explanation.
2. Compare and contrast feature generation and feature selection algorithms.
3. Discuss on machine learning and its types.
4. Discuss common methods of data validation that can be applied to ensure the quality and integrity of the dataset.
5. Briefly explain the ETL and ELT process of data migration.

**Group B**                      **[5 × 6 = 30]**

6. Elaborate on TDSP (Team Data Science Process) as a framework for the data science lifecycle.

**OR**

7. Discuss CRISP-DM (Cross-Industry Standard Process for Data Mining) as an agile approach to the data science lifecycle.
7. Consider a dataset representing whether students passed an exam based on three features: Study Hours (Low, Medium, High), Previous Grades (Low, Medium, High), and Tutoring (Yes or No). The target variable is Exam Result (Pass or Fail).

Study Hours	Previous Grades	Tutoring	Exam Result
Low	Low	Yes	Fail
Low	Medium	No	Fail
Medium	High	Yes	Pass
High	Low	No	Fail
Medium	Medium	Yes	Pass
High	High	Yes	Pass
High	High	No	Pass
Low	Low	No	Fail

Using the ID3 algorithm, calculate the information gain for each feature (Study Hours, Previous Grades, Tutoring) and determine which feature should be chosen as the root node for the decision tree.

**OR**

Consider a dataset containing the coordinates of 8 points in a two-dimensional space:

- Point 1: (2, 3)
- Point 2: (3, 4)
- Point 3: (3, 5)
- Point 4: (4, 6)
- Point 5: (7, 8)
- Point 6: (8, 7)
- Point 7: (9, 8)
- Point 8: (10, 9)

Apply the K-Means algorithm to cluster these points into 3 clusters.

8. You are analyzing a dataset containing information about customer orders for an e-commerce platform. However, upon initial inspection, you notice several data quality issues that may impact the reliability of your analysis.

Describe three common data quality issues that you may have identified in the dataset, providing specific examples for each issue. Explain the potential consequences of these issues on your analysis and propose strategies to address them effectively.

9. Explain the linear regression algorithm with appropriate example.

10. Consider a dataset containing monthly sales data for a retail store over a period of two years. The dataset consists of the following columns: Date (representing the month), Sales (the total sales for that month) and profit. Using this dataset, answer the following questions:

- a) Define what a time series is and explain its importance in data analysis.
- b) Identify and describe the different types of time series patterns that may exist in the sales data.

\*\*\*



Tribhuvan University  
Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
First Assessment 2080

**Subject: Database Management Systems**  
**Course No: MDS 505**  
**Level: MDS /I Year/I Semester**

**Full Marks: 45**  
**Pass Marks: 22.5**  
**Time: 2 hrs**

*Candidates are required to give their answer in their own words as far as practicable.*  
**Attempt ALL questions.**

**Group A [5 × 3 = 15]**

1. Differentiate network data model from hierarchical data model. [3]
2. Define relationship set and relationship instance. [3]
3. Why aliasing in SQL is needed. Show aliasing of a relation in SQL. [1+2]
4. Describe attribute, relation and tuple with examples. [3]
5. While converting ER model into relational schema, how the mapping of binary M:N relationship types is done? Illustrate with an example. [2+1]

**Group B [5 × 6 = 30]**

6. How tuple calculus is different from domain calculus? For following relation, write tuple calculus statement to retrieve name and roll number of all students who studies at school named MDS. [3+3]  
Student(Roll, Name, Address, Sid)      School(Sid, Sname, Affiliated\_to)

**OR**

How relational algebra is different from tuple calculus? For following relation, write domain calculus statement to retrieve name and roll number of all students who studies at school named MDS. [3+3]

Student(Roll, Name, Address, Sid)  
School(Sid, Sname, Affiliated\_to)

7. What is stored procedure? Mention its use. Given following relations, create a stored procedure named check person() to ensure that there is a person who is not a student. [1+1+4]  
Person(Pid, Pname, Page, Paddress)      Student(Roll, Sname, Sage, Pid)

**OR**

What is trigger? Mention its use. Given following relations, create a before insert trigger on person that will call checkperson() procedure when age of person is greater than age of student. [1+1+4]

Person(Pid, Pname, Page, Paddress)  
Student(Roll, Sname, Sage, Pid)

8. Design an ER diagram for following scenario; [6]  
In film industry, producers produce movies. Producers have their name, age and budget as attributes. They are uniquely identified by prod\_id. All the movies have their title, year, and release date. No movies can have same title. Every movies must be played by an actor. An actor can play many movies. Actors have Fname and Lname to uniquely identify them. The actors have charge\_rate as well. A single movie can have many producers and a producer can produce zero or many movies. There is an identifying relationship between actor and vanity van. Vanity van has partial attribute van\_id.
9. Consider the following relations for hospital management system; [6]  
Doctor(Did, Dname, Dspecialization)  
Patient(Pid, Pname, Paddress, Mname)  
Appointment(Did, Pid, Date, Time)  
Medicine(Mname, Expiry\_date, Composition)  
Write the **SQL and relational algebra** statements for following;  
a) Find name and address of all patients.  
b) Find name of doctors who have appointment on 2024/04/08.  
c) Find names of patients and doctors with whom patients have appointments.  
d) Find the names of patients and the names of medicines taken by them. Also list the number of medicine taken by the patients.
10. What is DBMS? Describe the characteristics that makes DBMS better than flat file system. [1+5]

\*\*\*

Tribhuvan University  
Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
First Assessment 2080

Subject: Statistical Computing with R  
Course No: MDS 503  
Level: MDS /I Year /I Semester

Full Marks: 45  
Pass Marks: 22.5  
Time: 2hrs

Candidates are required to write answers with examples for answering question numbers 1-5 in answer sheets and use R studio for answering question numbers 6-10 in your laptop. R scripts and knitted HTML files (outputs with interpretation) of question number 6-10 must be saved in a folder with name/exam roll number and submit them for grading.

Attempt ALL Questions.

**Group A** [5 × 3 = 15]

1. Explain how to import these types of data in R using base R functions:
  - a) Comma separated values text file
  - b) Excel data file
  - c) SPSS data file
2. Explain how you can do sub-setting with codes in R software:
  - a) Define the 6x5 matrix and select last two rows
  - b) Select third and fifth row with second and fourth column
  - c) Add 3 new rows in this matrix
3. Explain differences of these terms with examples using R codes:
  - a) Arrays and matrices
  - b) List and factors
  - c) Data frame and tibble
4. Explain the following concepts on working efficiently with "big data" in R software:
  - a) Sample and model
  - b) Chunk and pull
  - c) Push compute to data
5. Explain different types of pipe operators with R codes and examples:
  - a) Compound assignment operator
  - b) Tee operator
  - c) Exposition operator

**Group B** [5 × 6 = 30]

6. Do the followings with R script in R Studio:
  - a) Define a column vector X with numbers between 1 and 30
  - b) Define another column vector Y with cubes of X
  - c) Combine the two column vectors in a new data frame called DF
  - d) Get plot X and Y variables and decide which type of relationship is seen
  - e) Show the code to make this a linear relationship
7. Create a function and do as follows:
  - a) Define a function: "roll" of a fair "die" twice with random sampling with replacement as true
  - b) Get the first roll and interpret the result
  - c) Get the second roll and interpret the result
  - d) Get the third roll and interpret the result
  - e) Write a summary of the results obtained in the earlier steps with conclusion



8. Do the following in R Studio with tidy verse package using R Script to knit HTML output:
- Define a tibble having country, year, cases and population variables with 10 random data each
  - Transform this tibble to long format and interpret it carefully in terms of tidy data format
  - Transform the cases variable as log of cases (LnCase) and population variable as log of population (LnPop)
  - Create scatter plots of 1. Cases and population, 2. LnCase and population, 3. Cases and LnPop and 4. LnCase and LnPop
  - Show the four graphs in a single graph window

9. Use the "mtcars" dataset of tidiverse package and do as follows with R script to knit HTML output:

- Plot histogram of mpg variable and interpret it carefully
- Refine the histogram by filling the bars with "blue" color and changing number of bins to 10
- Add a vertical abline at mean of the mpg variable
- Plot Q-Q plot of mpg variable, add normal Q-Q line of red color on it and interpret it carefully
- Plot density plot of mpg variable without the border, fill it with yellow color and interpret it

OR

Use the "air quality" dataset of R to do following using base R to knit HTML output with R script:

- Create line plot of "Temp" with "Day" as the row index and interpret it carefully
- Create bar plot of "Temp" variable after defining class intervals systematically
- Create histogram of "Temp" variable and compare it with the bar plot of "Temp" variable
- Plot Normal Q-Q plot of "Temp" variable and interpret it carefully
- Create a scatter plot of "Temp" and "Wind" variables and interpret it carefully

10. Load the "igraph" package in R studio and do the basic SNA as follows with R scripts to knit HTML output:

- Define g as graph object with (1,2,3,4) as its elements
- Plot the g and interpret it carefully
- Define g1 as graph object with ("Sita", "Ram", "Rita", "Gita", "Gita", "Sita", "Sita", "Gita", "Anita", "Rita", "Ram", "Sita") as its elements
- Plot g1 with node color as green, node size as 20, link color as red and link size as 10 and interpret it
- Get degree, closeness and betweenness of g1 and interpret them carefully.

OR

Do as follows in R console and then to R Studio with R script to knit HTML outputs:

- Open R console and then go to Help and Manuals (in PDF) and open "An Introduction to R" file
- Save this file in the working directory and import this pdf file in R studio using "pdftools" package
- Perform pre-processing and create 'corpus' using "tm" package
- Find the most frequent terms and create histogram of the most frequent terms
- Create word cloud of the corpus, color it using rainbow or R Color Brewer package
- Perform topic modelling and interpret the result carefully

\*\*\*