

# DATA AND TEXT MINING

DOCUMENTATION RESEARCH AND TRAINING CENTRE, ISI-BC

MID-TERM EXAMINATION (2024)

Total Marks: 40

Time: 90 mins.

## ANSWER ALL

1. Explain the major challenges of traditional and human data analysis methods [3+3]
2. How do you measure the orthogonality and orthonormality of two vectors? [3+3]
3. How Eigen values and Eigen Vector help to make a better decision. [2+2]
4. What is the process of Knowledge Discovery in Databases? Describe each of its operations with a diagram. [4]
5. Differentiate between feature selection and feature extraction operations in a decision-making process. How they help to make better decisions. [4]
6. Find the variance-covariance matrix for the following dataset [6]

$XX =$

4	5	3
3	2	3
3	2	4
1	0	7
2	3	4
5	7	8
7	9	4
3	7	6
2	1	1
1	0	8

**OR**

Describe FOUR major types of attributes with examples, used in the data mining operation and provide the properties they possess. [6]

7. How positive definiteness property of the variance-covariance matrix is required in finding the dissimilarity measures between data points? Explain with an equation. [5]

**OR**

Explain the similarity measures between two vectors and their significance. How similarity measure is different from dissimilarity measures? [5]

8. Write THREE properties of matrix multiplication. [3]
9. Find the Euclidian norm of a vector  $X = [2 \ 4 \ 6 \ 3]$ . [2]

=====END of the Question Paper=====