

INDIAN STATISTICAL INSTITUTE

(MS-QMS)_2023

MID-SEMESTER EXAMINATION

(Pattern Recognition)

Duration: 90 minutes

Maximum Marks: 40

Note: Answer all question

1. Differentiate between NOMINAL, BINARY (Symmetric and Asymmetric) and Ordinal attributes with their mathematical properties. [3]
2. Describe four important information pre-processing steps with examples in a PR system. [4]
1. Name TWO significances of Eigen values and Eigen vectors in a database of 10 samples and 5 features for each sample. [2]
3. (a) What is interquartile range and find it for the following vector.
 $X=[12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25]$ [1+2]
(b) Check with interquartile range, if any outlier present in the above vector. [2]
4. What is the influence of non-negative-definite and positive-definite variance-covariance matrix on the PDF for normal distribution? Your answer should be supported with proper description and equation. [5]
5. Describe with examples and equations; the measures of Location, spread, Shape and dependency. [1x4]
6. Describe the motivations of standardization and normalization of data sets? Give at least one method of performing these operations [2+2]
1. If X is a data set with 9 samples and 2 features each. Find the Euclidean and Mahalanobis distances between 3rd and 5th samples of X. Which one is more significant and why?
 $XX = \begin{bmatrix} 4 & 5 \\ 3 & 2 \\ 3 & 2 \\ 1 & 6 \\ 0 & 2 \\ 5 & 7 \\ 7 & 9 \\ 3 & 7 \\ 2 & 1 \end{bmatrix}$; [6]
7. Differentiate between similarity and dissimilarity measures. Explain each measure with one method. [3]
8. Describe the method with equation of finding similarity and dissimilarity measures for Nominal and Ordinal attributes. [2]
9. What is positive definiteness property of distance metric? Explain with example [2]