

DATA AND TEXT MINING

DOCUMENTATION RESEARCH UNIT, ISI-BC

Total Marks: 30

Time: 1.5 hrs.

ANSWER ALL

1. Describe with examples and equations; the measures of Location, spread, Shape and dependency. [8]
2. Find and draw the BOX-and-WHISKER plot analysis for the following data set [2]
 $X=[4.3, 5.1, 3.9, 4.5, 4.4, 4.9, 5.0, 4.7, 4.1, 4.6, 4.4, 4.3, 4.8, 4.4, 4.2, 4.5, 4.4]$
3. Describe the motivations of standardization and normalization of data sets? Give one method for performing these operations. [2+2]
4. Define the relationship between Euclidean, normalised Euclidean and Mahalanobis distance with equations. [6]
5. List out the basic difference between Data Mining ~ Statistical Analysis [4]
6. [3]
 - a. What is knowledge discovery and database (KDD) process? Explain the structural view and the different components of the KDD system.
 - b. Mention the major issues in data mining. [3]