

Paper-21

# DATA AND TEXT MINING

DOCUMENTATION RESEARCH UNIT, ISI-BC

Total Marks: 30

Dt: 23/02/16

Time: 1.5 hrs.

## ANSWER ALL

1. Write short notes on different learning methods [3]
2. Explain. [2+2]
  - a. What is the motivation for data mining? Explain in contrast to the problems of conventional data analysis.
  - b. List at least FIVE application areas of data mining
3. Find and draw the BOX-and-WHISKER plot analysis for the following data set [3]  
 $X=[4.3, 5.1, 3.9, 4.5, 4.4, 4.9, 5.0, 4.7, 4.1, 4.6, 4.4, 4.3, 4.8, 4.4, 4.2, 4.5, 4.4]$
4. Describe the motivations of standardization and normalization of data sets? Give at least one method of performing these operations [2+2]
5. Define the relationship between Euclidean, normalised Euclidean and Mahalanobis distance with equations. [6]
6. Explain with examples for different types of attributes. [4]
7.
  - a. What is frequency, mean, mode, median and percentiles of a data sequence? [6]
  - b. What is the relation among mean, median and mode of symmetric, negatively and positively skewed data?