

Paper-2)

# DATA AND TEXT MINING

DOCUMENTATION RESEARCH UNIT, ISI-BC

## FINAL EXAMINATION

Total Marks: 50

Dt: 10/06/16

Time: 3 hrs.

1. Explain [4]
  - a. What is cross validation? And why it is required?
  - b. Discuss different methods of cross validation.
2. What is ROC curve and how it is constructed? [4]
3. Describe the motivations of standardization and normalization of data sets? Give at least one method of performing these operations. [4]
4. What is Bayes decision rule? Derive the mathematical formulation of classification errors with Bayes decision rule. [8]
5. Write the KNN algorithm. Describe its merits and demerits. [6]
6.
  - a. What is cluster analysis? Give the mathematical formulation of defining clusters [4]
  - b. Write the K-means and DBSCAN clustering algorithms with their advantage and disadvantages over each other. [4+4]
7. What is principal components? Describe the procedure to get principal components. [6]
8. Describe the Branch and Bound Feature selection method with an example of selecting TWO optimum features out of SIX features. [6]