

Paper: 21

DATA AND TEXT MINING

DOCUMENTATION RESEARCH UNIT, ISI-BC

Total Marks: 30

Dt: 24/02/16

Time: 1.5 hrs.

ANSWER ALL

1. What is variance-covariance matrix? Discuss the positive definiteness property of variance covariance matrix in the context of distance measure. [4]
2. Establish the relationship between variance-covariance matrix with correlation matrix. [2]
3. Why we go for data preprocessing? Discuss any FIVE preprocessing steps. [5]
4. Write short notes on different learning methods [6].
5. 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
6. Describe [2+2]
 - a. Differentiate between data mining and statistical analysis of data.
 - b. Describe the KDD process with figure. Show the importance of data mining.
7. Given the following confusion matrix of a classifier, find [5]
 - a. Accuracy
 - b. Precision
 - c. Recall
 - d. Sensitivity
 - e. Selectivity of the classes

	PREDICTED CLASS	
	Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	6954
	Class=No	412
		46
		2588