

Indian Institute of Engineering Science and Technology, Shibpur  
B.Tech (CST) 4 th Semester End Semester Examination, May 2021

Subject: Introduction to Data Science

Code: CS2205

Full Marks: 50

Time: One and half hour

**Answer question No. 5 and any two from the rest**

1. (a)  $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$

Find the Eigen values and singular values of Matrix A showing step by step decomposition method. 10

1.(b) Write the properties of Eigen vector. 7

1.(c) What is the limitation of Eigen vector decomposition method? 3

2.(a) Write the drawbacks of K-means clustering algorithm. 8

(b) SVM, Naïve Bayes and logistic regression classifiers are suitable for which kind of datasets and why? 4+8

3(a) What are the characteristics of social network? 5

(b) Explain Girvan-Newman (GN) Algorithm to calculate betweenness of an edge in a social graph. 15

4. (a) Why Named Entity Recognition is an important application in Text analytics? 5

(b) What are the metrics used to measure performance of a classifier? 12

(c) Why do you use clustering index? 3

**Answer any four from question No. 5**

5. (i) Why Kernel function is used in SVM classifier? 2.5 × 4

(ii) Write different steps for analysing an unstructured text document.

(iii) Why traditional clustering algorithms are not suitable for community detection in a social graph?

(iv) Write the purpose of using the parameters (minPts and esp) in DBSCAN algorithm.

(v) What is the condition to get orthogonal eigen vectors of a matrix A?