

## **CS4038D Data Mining**

### **Assignment No: 2**

**Due Date: 16.10.2020**

**Topic:** Implement Bisecting k-means clustering algorithm using a standard dataset and compare the performance with the previously implemented k-means algorithm.

- 1) Write a program to implement Bisecting k-means clustering algorithm by using Iris data set (available in UCI Machine learning repository) and find the followings:
  - i) Clusters of the Iris data set (final clustering solution).
  - ii) Sum of the Intra-Cluster Distances (SICD) or Sum of the Squared Error (SSE) values of the obtained clustering solution.
  - iii) Graphical representation of the obtained clusters.
  - iv) Give 500 runs of the implemented Bisecting k-means by changing the initial centroids and see the changes in the clustering solution (SICD values).
  - v) Compare k-means and Bisecting k-means with respect to iteration vs. SICD plot.

**Students are requested to take care of the following points:**

- 1) Download the dataset from UCI repository, do a shuffling of the records/tuples/rows/objects/data points of the dataset such that same class data points or objects are not placed one after another and then remove the class label attribute.
- 2) Please don't use inbuilt functions available in python or in some other platforms for distance measures and coding etc.
- 3) Save your solution document as a .pdf file. (Format: "Reg.NumberYour NameExercise No.pdf" (Ex: B190585CS JohnE01.pdf).
- 4) Submit your pdf file to Eduserver before deadline (**16.10.2020, 23.59 PM**).
- 5) Late submission may incur penalty of **-0.5** marks.
- 6) Evaluation will not be conducted for those who don't submit their solution on time.
- 5) Viva (Interactions) will be conducted through WebEx/Google meet. (Share your screen and show your execution)

- 6) Due to unpredictable situations, students who are not able to appear for online viva sessions, may get chance to attend the oral viva through mobile phone.

Note: Only genuine cases will be entertained and students will have to get prior permission from me for the same.

- 7) WebEx join meeting link details will be sent to the students one day earlier.
- 8) You may use any programming language of your choice.
- 9) Please take snapshots of all your executions and put them in a word file and save it as a .pdf file for submission. If any doubts on the uploaded marks and uploaded exercises, please mail me (praneshdas@nitc.ac.in).

\*\*\*\*\*