Assignment 2: DSML Course (DSE 302/ECS 308/ECS 658)

Indian Institute of Science Education and Research Bhopal

Deadline: April 16, 2022, 23:59 IST **Full Marks: 10**

Overview

You have to work on an unsupervised problem using a given **synthetic data set**. The dataset is uploaded in this **LINK**. **The true labels (i.e., ground truths) of the given data points will be released after the deadline**. The objective is to develop or propose an effective unsupervised technique that can identify the natural groups (or clusters) in the given data set.

Tasks

You have to perform the following tasks using the given data.

1. Plot the given data points on the console in order to visualize it. Your code must reproduce the plots on re-running it.

   [2]

2. Develop and then implement an effective unsupervised method/framework that can perfectly find the natural clusters of the given data set. **Assume number of clusters is 2**. You must explore the performance of different relevant feature engineering and clustering techniques in order to achieve this.

   [6]

3. Submit the predicted cluster labels (by the method/framework developed) of the given data points in a text file. The performance of your model will be evaluated in terms of normalized mutual information following this **library**, but you cannot test it before the deadline. Your code must reproduce the cluster labels which will be uploaded.

   [2]

Submission Guidelines

The cluster labels of the individual data points must be written in a text file **in the order the data points are given**. Any two cluster labels must be separated by a newline in this text file. **Any other style of the text file will not be accepted and it will be graded as 0.** This text file along with the code must be uploaded in the given Google classroom link for this assignment. The code must reproduce the same result as reported in the text file by running it on the given data.

Other Relevant Information

- Multiple submissions are allowed within the deadline, but only the last submission will be graded. You must upload the submission in the Google classroom. In case of any technical difficulty to upload it to Google classroom, you may drop an email to tanmay@iiserb.ac.in stating the same, much before the deadline.
- **For every half an hour that your submission is late your score gets multiplied by 0.8 till it becomes 1.**