Assignment 1: DSML Course (DSE 302/ECS 308/ECS 658)

Indian Institute of Science Education and Research Bhopal

Deadline: February 27, 2022, 23:59 IST **Full Marks: 10**

Overview

You have to work on a binary classification problem on a given **synthetic data set**. You have to explore the performance of different classification techniques as discussed in the class. The training data along with the class labels and the test data are uploaded in this **LINK**. The class labels are marked as 0 or 1. The objective is to train a classifier on the training data and find it's best setting and then use this setting to classify the instances of the given test data. **The class labels of the test data will be released after the deadline**.

Tasks

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

1. Plot the given training data using different colors for individual classes on the console in order to visualize it. Your code must show this plot on re-running it.

   [2]

2. Perform naive bayes, logistic regression and k-nearest neighbor classifiers to categorize the instances of the given test data. Initially, find the best performance for each of these classifiers by properly tuning its parameters as discussed in the class.

   [4]

3. Submit the class labels of the test data instances in a text file. The performance of your model will be evaluated in terms the macro-averaged f-measure following this **library**. Your code must reproduce these results and **must be uploaded**.

   [4]

Submission Guidelines

The class labels (0 or 1) of the test data instances must be written in a text file **in the order the instances of the test data instances are given**. Any two class labels must be separated by a newline in this text file. The pattern of this text file will be same as the file containing training data class labels. Any other style will not accepted and it will be graded as 0. This **text file along with the code** must be uploaded in the classroom link for this assignment. The code should print the performance of the best setting of all the classifiers on the training data following this **library**.

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.**