Assignment: Machine Learning (DSE 317/617 & ECS 317)

Indian Institute of Science Education and Research Bhopal

Deadline: October 07, 2023, 23:59 IST **Full Marks: 10**

Overview

The objective is to work on an classification problem using a given **synthetic data set**. The dataset is uploaded in this **LINK**. The training data along with the class labels and the test data are uploaded in the link. **The actual class labels (i.e., ground truths) of the given test data points will be released after the deadline**. The objective is to explore the existing techniques or propose an effective classification technique that can perfectly classify the given data set. The rows of the given data file indicate instances and columns indicate features.

Tasks

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

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

   [1]

2. Explore the best performance of Decision Tree, Random Forest, k-Nearest Neighbor and Logistic Regression classifiers individually using the training data. Use **grid search** method to tune the parameters of individual classifiers. Report the precision, recall, specificity and f-measure (i.e., f1-score) of these classifiers in a table and then analyze the results. Implement the best classifier (as per the training data) on the test data set and submit the class labels in a separate text file. Your code must reproduce the class labels that you will upload.

   [4]

2. Write the code for precision, recall, specificity and f-measure on your own. You cannot use scikit-learn or any other standard library for the same.

   [2]

3. Suggest and implement a feature transformation technique to perfectly classify all the data points using the best classifier in point 2. Submit the class labels in another text file. Your code must reproduce the class labels that you will upload.

   [3]

Submission Guidelines

The class labels of the individual data points must be written in a text file **in the order the data points are given**. Any two class 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 0.**