Lab Assignment-6

CSL2010: Introduction To Machine Learning

Support Vector Machines

AY 2021-22, Semester-I

General Instructions:

1. Prepare Python code files and name them <ROLL_NO>.py
2. Also, provide your Colab file link in the report. Make sure that the file is shareable.
3. Submit a single report depicting the method, results, and observations for all the tasks. There is no need to add theory behind the concepts. Name the report as <ROLL_NO>_Report.pdf
4. Clearly, mention the assumptions you have made, if any.
5. You are free to use any library.
6. Clearly, report any resources you have used while attempting the assignment.
7. Any submission received in another format or after the deadline will not be Evaluated.
8. Task 1 will be a separate assignment in the google classroom and Task 2 and Task 3 will be together in another classroom
9. For Task 1 in Google Classroom just give one file, <ROLL_NO>_Task1.py
10. For Task 2 in Google Classroom just give two files, <ROLL_NO>_Task2.py, and <ROLL_NO>_Report.pdf
11. Note: Plagiarism of any kind will not be tolerated and will result in zero marks.

Assignment

1. **Task1 (5 marks) (Deadline: 16 Sep 2021, 11:59 PM)**
    a. Download the dataset from the below link https://drive.google.com/file/d/1tsZLLpcw6BWKmb1Jqym88KLEmnYdEchJ/view?usp=sharing
    b. Store the dataset in your google drive and in Colab file load the dataset from your drive
    c. Check the shape and head of the dataset
    d. Identify the features and targets from the data and take into a separate variable
    e. Split the dataset into 80:20.
    f. Train the model using scikit learn SVM API (LinearSVC with C as 1.0)
    g. Print score on test data
    h. Make the prediction on test data
    i. Print confusion matrix and classification report

2. **Task 2 (15 marks) (Deadline: 22 Sep 2021, 5:30 PM)**
    a. Download the dataset from the below link https://drive.google.com/file/d/1XDHNRbTui9QzjkObm_G0_A6My9QIb05v/view?usp=sharing
    b. Store the dataset in your google drive and in Colab file load the dataset from your drive
    c. Check the shape and head of the dataset
    d. Age, Experience, Income, CCAvg, Mortgage, Securities are the features and Creditcard is your Target Variable
        i. Take any 3 features from the six features given above
        ii. Store features and targets into a separate variable
        iii. Look for missing values in the data, if any, and address them accordingly.
        iv. Plot a 3D scatter plot using Matplotlib.
    e. Split the dataset into 80:20. (3 features and 1 target variable)
    f. Train the model using scikit learn SVM API (LinearSVC) by setting the regularization parameter C as C={0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000}
        i. For each value of C Print the score on test data
        ii. Make the prediction on test data
        iii. Print confusion matrix and classification report
    g. Use gridSearchCV a cross-validation technique to find the best regularization parameters (i.e.: the best value of C)

Note: In the report provide your findings for the output generated for all the kernels used and also describe the changes that happened after changing the regularization hyperparameter.