Assignment 06

The following instructions are step by step procedures of one single task

1. Collect the dataset you received in Lab 5, load it using pandas

2. Apply necessary pre-processing steps on it

3. Support Vector Machine (SVM), Neural Network (Multilayer Perceptron Classifier) and Random Forest are three very popular machine learning classifiers. Divide the dataset into 8:2 train-test split and perform Support Vector Machine, Neural Network (MLPClassifier) and Random Forest on it using sklearn library. In the previous assignment, you have already used Logistic Regression and decision tree classifiers from the sklearn library. Just change the imports and the function calls to use other classifiers. Take a look at the sklearn documentation for further information.

4. Perform dimensionality reduction using PCA. Reduce the number of feature vectors into half (e.g. if your dataset has 10 columns, after applying PCA it should have 5 columns)

5.Apply Support Vector Machine, Neural Network (MLPClassifier) and Random Forest again on the reduced dataset.

6. Compare the accuracy of the pre-PCA and post-PCA results. Plot them against each other in a bar graph.

7. Copy all your code in a doc file and submit the doc file

For sklearn documentation, you may visit this link