Assignment 4
Machine Learning (CS564)

Date: 13-Nov-2020 **Deadline:20-Nov-2020**

Instructions:
1. You are supposed to write code from scratch for methods but you can use libraries for pre-processing and post-processing.
2. The assignments should be completed and uploaded before the deadline.
3. Markings will be based on the correctness and soundness of the outputs. Marks will be deducted in case of plagiarism.
4. Proper indentation and appropriate comments are mandatory.
5. You should zip all the files and name the zip file as roll_no.zip, eg. 1701cs11.zip.
6. Make necessary assumptions if required. For further clarification, you can write an email: cs5642020@gmail.com

Upload your zip file to the link: https://www.dropbox.com/request/S2GoAO6C3S6qE5tsiplS
Dataset:
https://www.dropbox.com/s/xuijoj5a34outec/ReplicatedAcousticFeatures-ParkinsonDatabase.csv?dl=0
Dataset Information:
https://archive.ics.uci.edu/ml/datasets/Parkinson+Dataset+with+replicated+acoustic+features+

PROBLEM

In this assignment you have to implement an Ensemble Classifier of Logistic regression, Naive Bayes & Decision Tree. The Goal is to implement a classification model to predict the "Status" feature in the provided Parkinson dataset.(Make sure to exclude the ID column). Make necessary assumptions regarding hyperparameters. Construct the following -

1. Ensemble using Majority voting ;
2. Ensemble using weighted voting; accuracy of the classifiers on the development set (20% of the overall data) is to be used as the weight.

A. For each of the ensembles compare the performance of the ensemble to each of the base models. Is there always a benefit of ensembles?

B. Redo ensemble with 5 versions of each of the specified classifiers(each version has different hyperparameters, if possible, for ex. For DT different versions can be generated by varying tree depth and branching factor, for logistic regression you can make variations by setting regularization etc.). How does this ensemble compare to the one with only one of each.