**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; the 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?

**Answer:** Seeing the results, we can conclude that the ensemble technique is working better for this composition and dataset. But, after performing some different compositions in the same experiment, I found that sometimes individual base models perform better. So, there is not always benefit from the ensembles and also there will be time and space constraints.

**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?

**Answer:** From the results, it can be clearly seen that Logistic regression with L1 and elastic net regularization doesn't work well. So, the results can be compared for L2 and no regularization. For L2- regularization, the ensemble methods are giving mixed results(sometimes better and sometimes worse) on changing the max-depths. But, with no regularization, ensemble methods are giving bad results.

The responses are as follows:

For decision tree depth 1:

Logistic regression using L1- Regularization:

Accuracy: nan (+/- nan) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68812 (+/- 0.06026) [Decision Tree]

Accuracy: nan (+/- nan) [Majority Voting]

Accuracy: nan (+/- nan) [Weighted Voting]

Logistic regression using L2- Regularization:

Accuracy: 0.75520 (+/- 0.07055) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68812 (+/- 0.06026) [Decision Tree]

Accuracy: 0.80216 (+/- 0.03074) [Majority Voting]

Accuracy: 0.79204 (+/- 0.05998) [Weighted Voting]

Logistic regression using elasticnet- Regularization:

Accuracy: nan (+/- nan) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68812 (+/- 0.06026) [Decision Tree]

Accuracy: nan (+/- nan) [Majority Voting]

Accuracy: nan (+/- nan) [Weighted Voting]

Logistic regression using no Regularization:

Accuracy: 0.74507 (+/- 0.07315) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68812 (+/- 0.06026) [Decision Tree]

Accuracy: 0.79190 (+/- 0.04518) [Majority Voting]

Accuracy: 0.80756 (+/- 0.06623) [Weighted Voting]

For decision tree depth 2:

Logistic regression using L1- Regularization:

Accuracy: nan (+/- nan) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68785 (+/- 0.03337) [Decision Tree]

Accuracy: nan (+/- nan) [Majority Voting]

Accuracy: nan (+/- nan) [Weighted Voting]

Logistic regression using L2- Regularization:

Accuracy: 0.75520 (+/- 0.07055) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68785 (+/- 0.03337) [Decision Tree]

Accuracy: 0.79703 (+/- 0.04996) [Majority Voting]

Accuracy: 0.79730 (+/- 0.05593) [Weighted Voting]

Logistic regression using elasticnet- Regularization:

Accuracy: nan (+/- nan) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68785 (+/- 0.03337) [Decision Tree]

Accuracy: nan (+/- nan) [Majority Voting]

Accuracy: nan (+/- nan) [Weighted Voting]

Logistic regression using no Regularization:

Accuracy: 0.74507 (+/- 0.07315) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.68785 (+/- 0.03337) [Decision Tree]

Accuracy: 0.79190 (+/- 0.05612) [Majority Voting]

Accuracy: 0.78138 (+/- 0.06636) [Weighted Voting]

For decision tree depth 3:

Logistic regression using L1- Regularization:

Accuracy: nan (+/- nan) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.72969 (+/- 0.06001) [Decision Tree]

Accuracy: nan (+/- nan) [Majority Voting]

Accuracy: nan (+/- nan) [Weighted Voting]

Logistic regression using L2- Regularization:

Accuracy: 0.75520 (+/- 0.07055) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.72443 (+/- 0.05796) [Decision Tree]

Accuracy: 0.78664 (+/- 0.04359) [Majority Voting]

Accuracy: 0.79730 (+/- 0.05593) [Weighted Voting]

Logistic regression using elasticnet- Regularization:

Accuracy: nan (+/- nan) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.73482 (+/- 0.05034) [Decision Tree]

Accuracy: nan (+/- nan) [Majority Voting]

Accuracy: nan (+/- nan) [Weighted Voting]

Logistic regression using no Regularization:

Accuracy: 0.74507 (+/- 0.07315) [Logistic Regression]

Accuracy: 0.79217 (+/- 0.06185) [Naive Bayes]

Accuracy: 0.72969 (+/- 0.06001) [Decision Tree]

Accuracy: 0.78165 (+/- 0.06084) [Majority Voting]

Accuracy: 0.77638 (+/- 0.06108) [Weighted Voting]