ADABOOST USING DECISION STUMPS

Username	Vikramaditya

Objective:

Implement Adaboost algorithm with Gini Index based Decision tree of depth equals one(decision stump).

Readings:

No. of Stumps	Training Error via self Implementation of AdaBoost with Decision Tree Classifier	Test Error(from miner) via self Implementation of AdaBoost with Decision Tree Classifier	Test Error via Lib Implementation of Decision Tree Classifier without Pruning
10	0.27	0.18	0.09
20	0.39	0.17	0.09
30	0.17	0.13	0.09
40	0.13	0.12	0.09
50	0.09	0.09	0.09
60	0.07	0.08	0.09
70	0.07	0.10	0.09
80	0.07	0.09	0.09
90	0.07	0.08	0.09
100	0.05	0.07	0.09
110	0.04	0.08	0.09
120	0.04	0.07	0.09
130	0.04	0.06	0.09
140	0.03	0.06	0.09
150	0.03	0.06	0.09
160	0.02	0.06	0.09
170	0.02	0.06	0.09
180	0.01	0.06	0.09
190	0.01	0.06	0.09
200	0.00	0.06	0.09
210	0.00	0.07	0.09
220	0.00	0.07	0.09

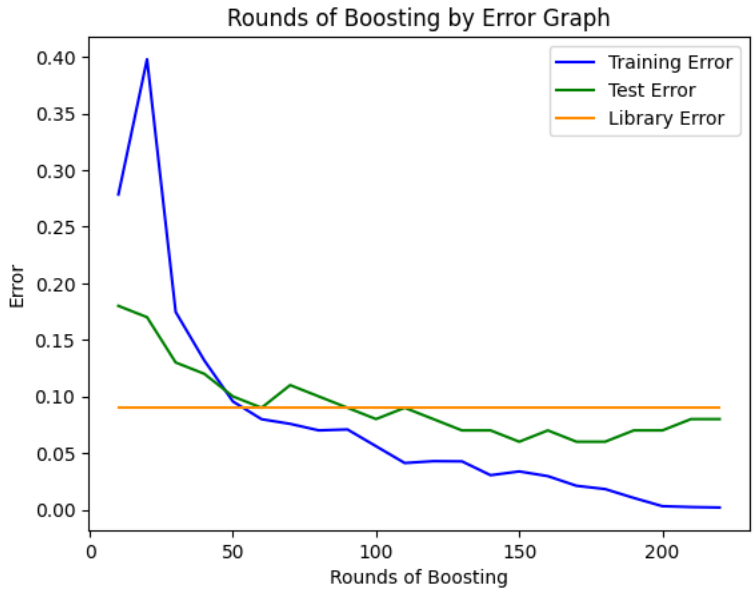
Observation and Inference

Following are the Observations:

- Training error was high when the model was trained using the initial number of stumps (up to 40 rounds), but when the number of stumps were increased further(from 200 to 220 rounds), we could see that training error was still decreasing and testing error was increasing, which shows overfitting.
- This shows that initially the model was completely unknown about the classes in the test set and therefore predicted the wrong response initially, but as the number of rounds of adaboosting was increased, the classifier gradually provided the right output prediction.

- This demonstrates that, for 150 to 190 classifiers, the Adaboost method would fit the training dataset properly. However, when the number of boosting rounds is increased, the algorithm would attempt to fit more classifiers to the training data, which will ultimately lead to overfitting.
- The testing error is least during 150 to 190 rounds of adaboosting, where it also exceeded the library implementation for decision tree without pruning.
- Sklearn library's test error utilizing a single decision tree without pruning provides an accuracy of 0.91 and a minimal error of 0.09.
- I tried the model with the ensemble size of 600, which provided me the accuracy of 0.90. Although I want to calculate the prediction for nearby to that ensemble size but the process of creation of 600 or more ensembles took lots of time thus making very difficult to record readings from miner.
- The Adaboost classifier predicted better above 90 rounds then a single Decision tree without pruning classifier. Therefore on plotting Receiving Operating Characteristic (ROC) curve graph for Adaboost classifier, the Area Under the Curve (AUC) will be greater for Adaboost in comparison to single Decision Tree. So Adaboost with Decision Stump is better.

Graphical Representation:



We can see from the graph above that the training error is highest in the first round of boosting, then it drops below 0.04 (4%), then drops to zero after 110 rounds, and finally to zero after 200 rounds. Overfitting is evident by this.

The estimated test error is at its greatest when beginning to increase and logically decreases with the training set error.

The orange horizontal line is the test error calculated by a Single Decision Tree trained using the Sci-kit Learn Library. This result from the built-in libraries exhibits consistency.

Conclusion:

As a result of the above, I was able to develop the Adaboost method with Gini Index-based Decision trees of depth one (also known as "decision stumps") as weak classifiers and test the model effectively on an unknown test dataset. A decision tree without pruning is also used, and the results are compared.