Victoria Capobianco
Dr. Brahma
Intro to Machine Learning
3/15/21

CA04 Written Response

1. Write your response about the Classifier's behavior with respect to the number of estimators.

In the Random Forest Classifier, the model's accuracy was decreasing steeply between 50-249 estimators. There was a sharp increase in accuracy between 250-300 estimators, and then a steady decline in accuracy with few variations betweens 301-450 estimators. After the 500 estimators point, the accuracy score is above where it was at the 300 point, but because we do not have any more data, we cannot say whether it continues to increase or decrease after this point. With this Classifier, there is clearly a threshold between too many (greater than 301) and not enough (less than 250) estimators which all result in a smaller accuracy score.

In the AdaBoost Classifier, the models' accuracy scores sharply increase between 50-100 estimators and then plateau between 100-150. After 150, the score sharply declines between 150-200 and then plateaus at a medium score of 0.844800 between 200-500. With AdaBoost, there is a clear range between 100-150 estimators for the classifier to behave the most accurately.

In the GradientBoost Classifier, there is a lot of variety in the accuracy scores across the different estimator values. Between 50-100 there is a sharp decrease in accuracy as the estimators increase. Then, between 100-200, there is a small and steady increase in accuracy. After that point, the accuracy score increases and decreases in intervals of 50 with the highest point at 350 estimators. GradientBoost is not consistent overall in its accuracy score given the estimator values.

Lastly, for the XGB Classifier, the accuracy score plateaued at 0.83 for every estimator value. Here, the number of estimators does not make a difference on accuracy score.

2. Is there an optimal value of the estimator within the given range?

No, there is not an optimal value that would satisfy every Classifier within the given range. The behavior of each Classifier varies even though they were given the same list of n_estimators. For example, in RandomForest Classifier one could say the optimal estimator is 500 because that was the highest score for that Classifier, but that was also a low or plateaued point for the other three Classifiers.