Carla Mariana Fera

CA04 – Ensemble Methods

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

   For each of the models, a different number of estimators is optimal. The optimal value of n_estimators for Gradient Boosting and XGBoost is higher than in AdaBoost or Random Forest. In addition, Random Forest doesn't seem to have a trend in what respects to n_estimators. As shown in the graph, the accuracy goes up and down for different number of estimators. In contrast, Adaboost Classifier's accuracy seems to decrease as the number of estimators increase. On the other hand, Gradient Boosting and XGBoost's accuracy increases as n_estimators go up.

   In what respects to the last table of the assignment, I assumed the same n_estimators for all algorithms. Thus, we can compare accuracy for the same level of estimators. As we can see, accuracy is lowest for Random Forest and increases for the other algorithms. The highest accuracy is for XGBoost. In addition, AUC is higher in Random Forest than Adaboost. However, Gradient Boosting and XGBoost have a higher AUC. If we compare this last two, Gradient Boosting AUC's is higher.

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

   For Random Forest, the optimal value is 150; for AdaBoost, the optimal value is 100; for Gradient Boosting, the optimal value is 250; and for XGBoost, the optimal value is 300.