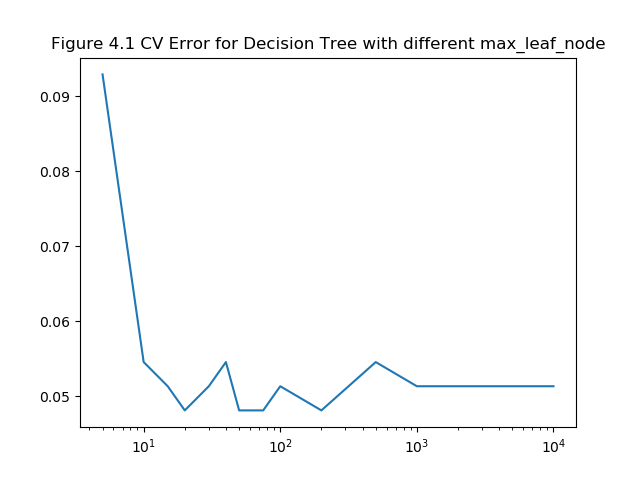
Homework 4: Decision Trees + Analysis

CS412

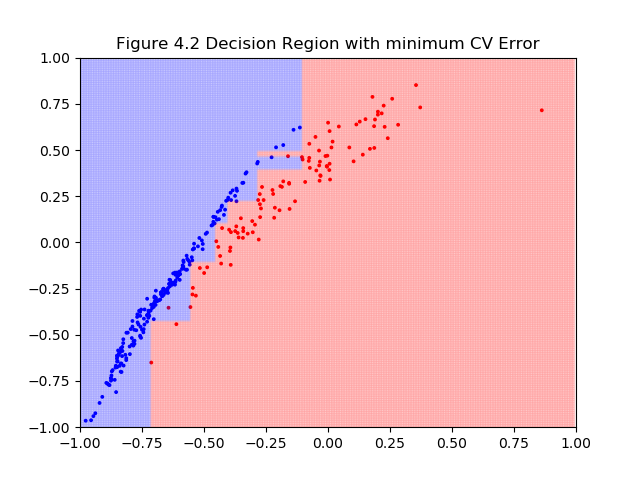
Due: April 23rd, 11:30pm on Gradescope

1 Trees/Ensemble Methods

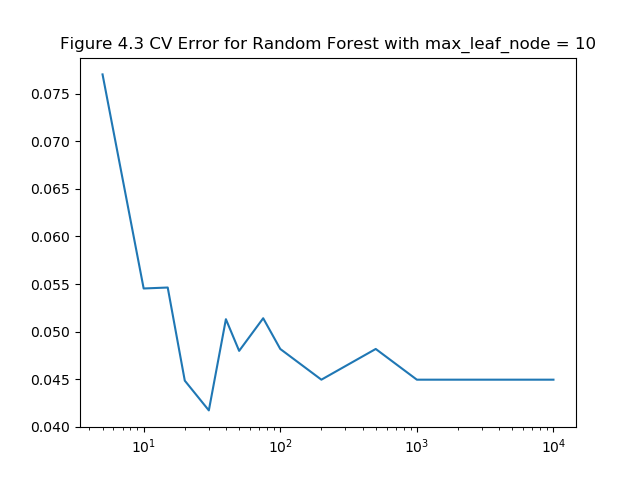
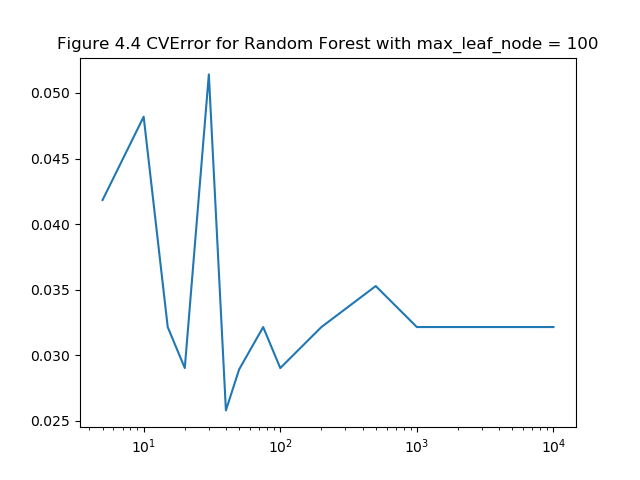


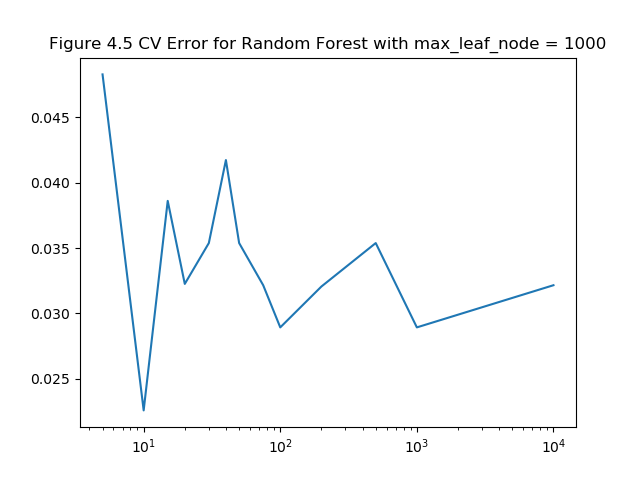


1. Looking at the graph, it looks like cv error below max\_leaf\_nodes = 15 is underfit since the CV error is too high for this trained model.



(e)

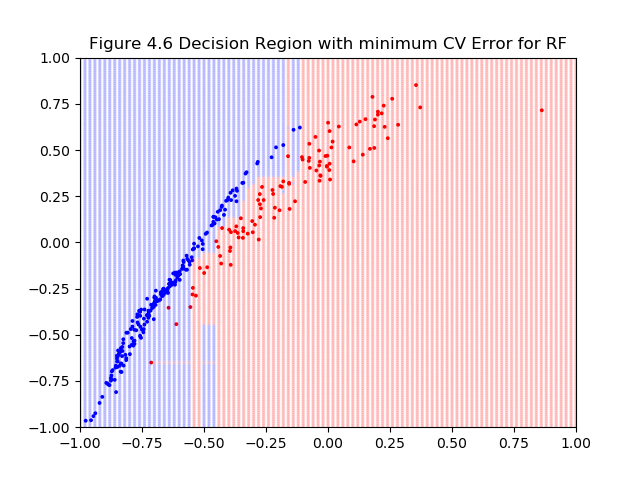




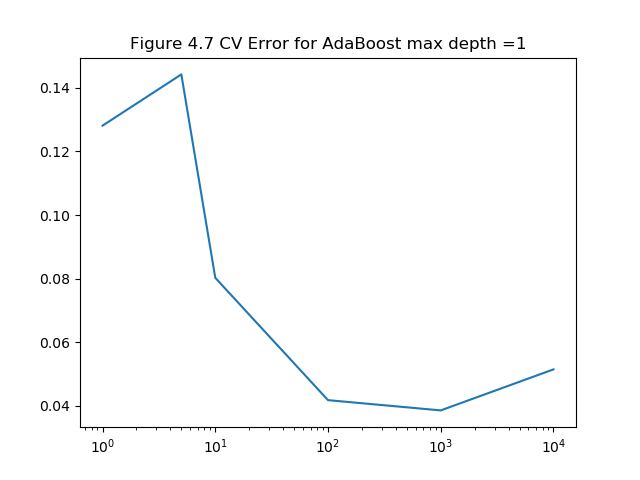
(f) According to the figure 4.5, max\_leaf\_nodes = 1000 was impacted the most since using the bagging approach on a well-trained model makes it overfit and hence the error starts going up slowly.

(g) N estimators do make impact on the result; it helps in increasing the accuracy of the model by increasing the number of trees in the forest. But after a point it stops impacting that much, in this case it is n\_estimator = 1000. Looking at the figure 4.3,4.4,4.5 we can observe that the error becomes constant even after increasing the n\_estimator. This also helps in knowing that now the model is getting overfit.

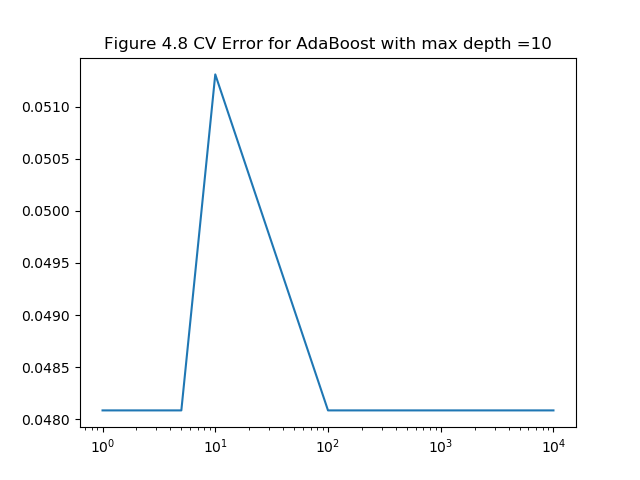
(h)



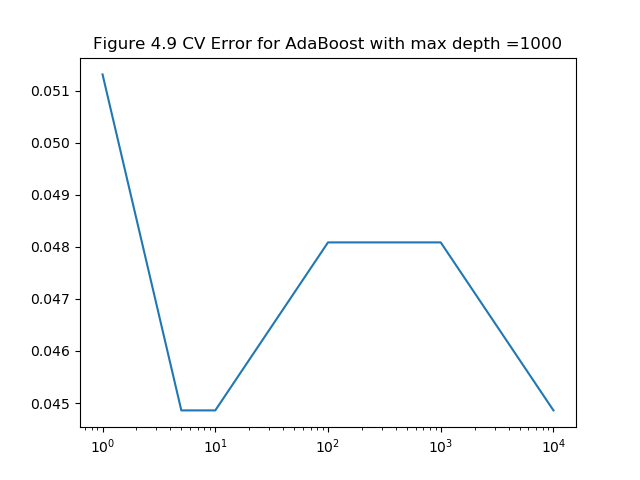
(j)



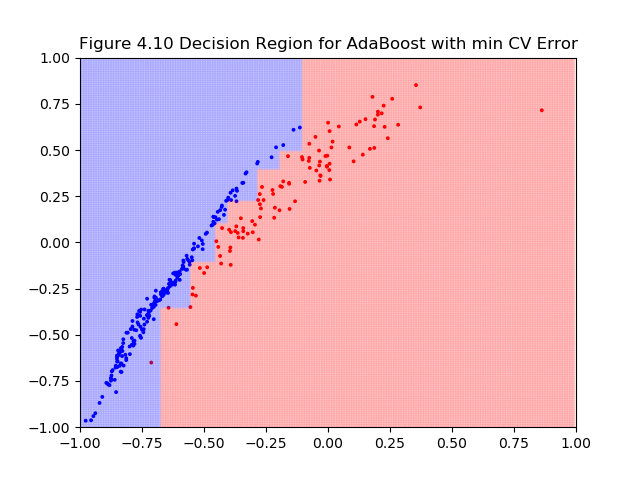
(k)



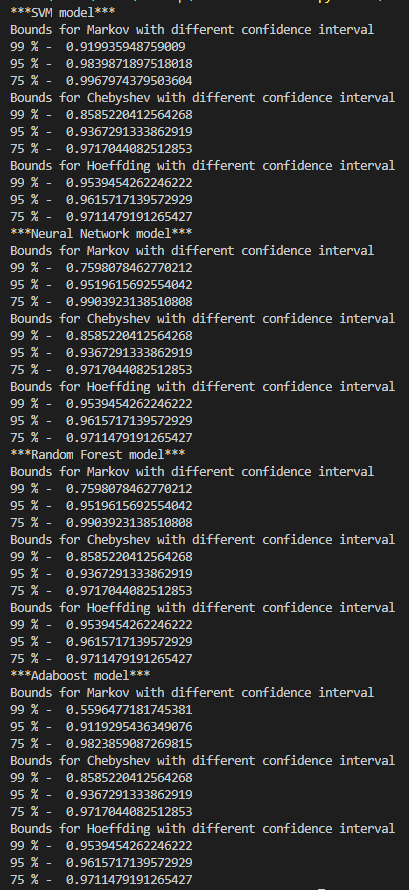
(l)



(m)



2 Reporting Final Error



1. According to the values, Adaboost is the model which changes drastically over the confidence interval
2. I would be choosing Neural Network as my model since it’s the most stable ML model in terms of the Final Errors.
3. The consideration I made for deciding is the sensitivity the model showed for hyperparameters. The model which doesn’t drastically changes by slightly tuning the hyperparameters are good models to be trained. As this way the model does not overfit and its easier to find that sweet spot between overfitted and underfitted model.