417 hw4

alex.huang

April 2020

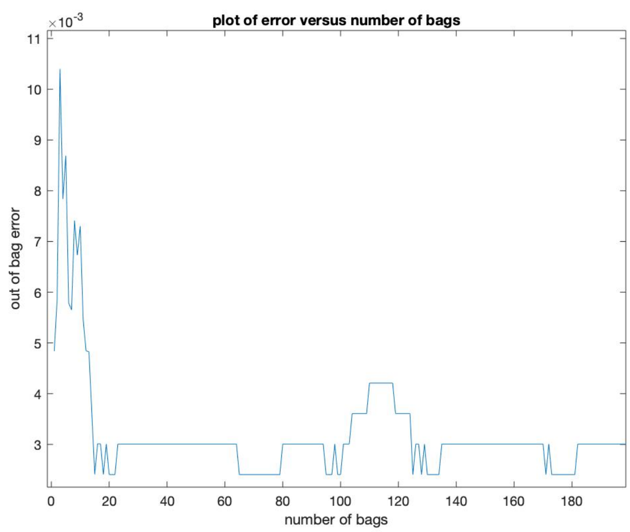
1

1.1

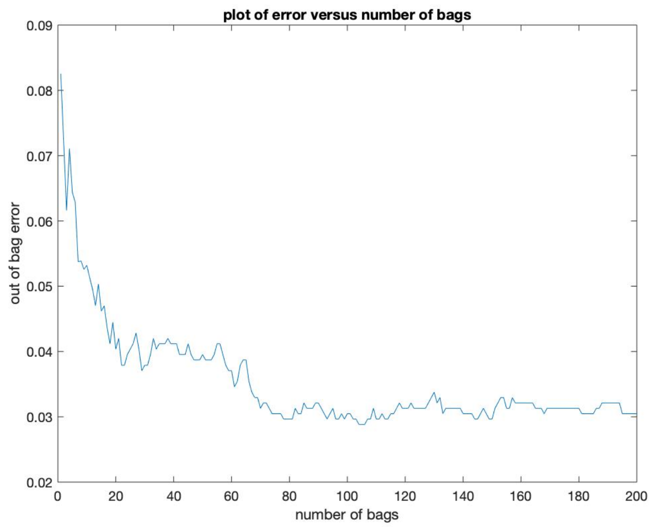
1.2

After running the script, we have the following result as the output:

The cross-validation error of decision trees is (for 1-3): 0.0066
The OOB error of 200 bagged decision trees is (for 1-3): 0.0030
Plot below



The cross-validation error of decision trees is (for 3-5): 0.0593
The OOB error of 200 bagged decision trees is (for 3-5): 0.0305
Plot below:



1.3

For one-versus-three problem, we have the test error for one decision tree tree and an ensemble of 200 to be 0.0163 and 0.0116, respectively.

For three-versus-five problem, we have the test error for one decision tree tree and an ensemble of 200 to be 0.1196 and 0.0828, respectively.

1.4

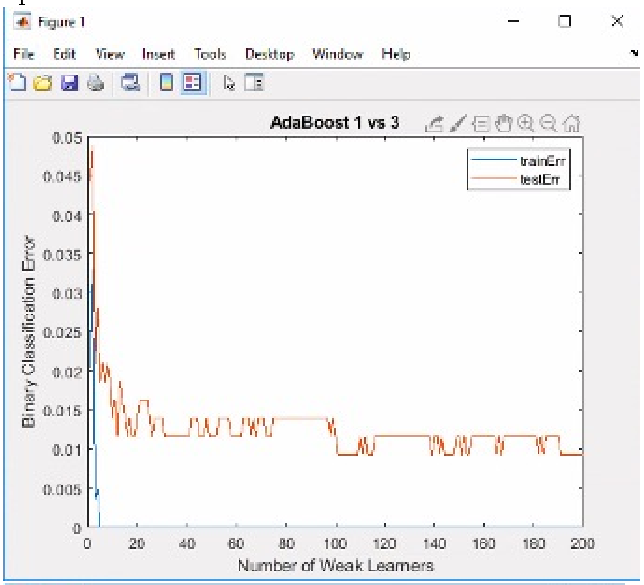
For all the data set, we can see that the OOB error of 200 bagged decision trees are always significantly better than the cross-validation error of decision trees. Also, we see that the test error is also better for an ensemble of 200. Therefore, we conclude that data suggest that bagging returns a better result (less out of sample error) than learning one decision tree.

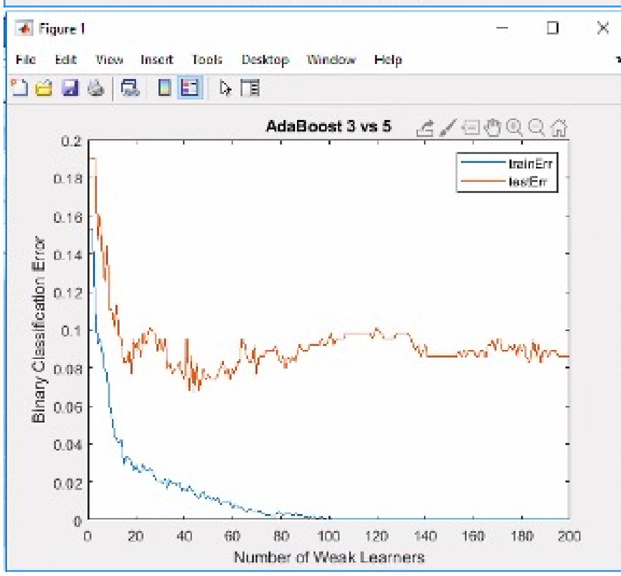
Also, if we compare the results generated by part b and c, we see that both the cross validation error and the OOB error is significantly lower than the test error. Therefore, we can conclude that both those errors are not the best estimates for out of sample error, possibly because their data points are selected from the learning data set, not the actual distribution. Therefore, we can use them for validation, in some occasions, but perhaps not for a compelling estimate of out of sample error. This also confirms our previous idea that validation errors are often optimistic estimates.

From the plots, we can see that the OOB error as a result of the learning first fell sharply, then stayed at a certain level. This suggests that as numBags gets larger, adding more bags will not significantly change our resulted effect of learning.

2

The pictures attached below:





In both graphs, we could see that in general both training error and test error decrease as the number of weak learners increases. The testing set error for 1-vs-3 problem drops to less than 0.02 after we have sufficient amount of weak learners (more than 10), and the testing error for 3 vs 5 problem drops to around 0.1 after more than 30 weak learners, which indicates that the performance of AdaBoost is pretty good for those two problems. However, depending on the specific problem we have, the number of weak learners that we need to make the training set error testing set error converge may vary. In particular, for the 1-vs 3-problem, the training set error drops to 0 for more than 5 weak learners. However, for the 3-vs 5-problem, it needs more than 100 weak learners to make the training set error 0.

Also worth noticing is the fact that the test error stays at basically the same level after dropping initially, indicating that after a certain number, increasing the number of weak hypotheses perhaps does not significantly change the out of sample error. Another trend is that the test set error can be consistently smaller than the training set error, which indicates that error in the training set is an optimistic estimate.