Homework 4 (Due Wednesday 4/10)
CS534 Machine Learning, Spring 2019

This homework will explore bagging and boosting. You will implement Adaboost.M1 and random forest algorithms, and compare your implementations to built-in implementations in a library of your choice.

Problem 1 - Adaboost M.1. (30 points)

Implement the Adaboost M.1 framework described in Algorithm Table 10.1 from your textbook. Use a built-in CART classification tree generator to create the weak learners (in Matlab this is *fitctree*). Pay careful attention to the tree generator parameter documentation. For each approach below, display the mean classification accuracy as a function of the weak learners or trees as directed. All problems will use the UCI ionosphere dataset.

1.a. Cross validation - tree size. Using 5 fold cross validation, compare boosting with stumps, 5-node trees, and 10-node trees. Generate classification accuracy versus the number of weak learners (up to $M = 300$) in each case. Compare each of these in a plot.

1.b. Cross validation - shrinkage. Apply *shrinkage* $\nu = 0.9$ for training the 5-node trees from above. Compare the shrunk and ordinary classifiers in a plot.

Problem 2 - Random forests (30 points)

Implement the random forest algorithm described in Algorithm Table 15.1 from your textbook.

2.a. Cross validation - m. Using the same cross-validation splits as above, compare random forests with $m = \{1, \sqrt{p}, p\}$. Allow the trees to grow deep by setting the maximum number of splits to $n - 1$ (where n is the size of the cross validation training samples). Compare these in a plot of classification accuracy versus number of trees.

2.b. Cross validation - n_{min}. Using the same splits and setting $m = \sqrt{p}$, compare random forests with and without depth control. To impose depth

control, set the min number of samples per terminal node $n_{min} = \{1, 10\}$. Choosing $n_{min} = 1$ does not constrain tree depth. Compare the depth control and ordinary classifiers in a plot.