CS760 Spring 2019 Homework 4
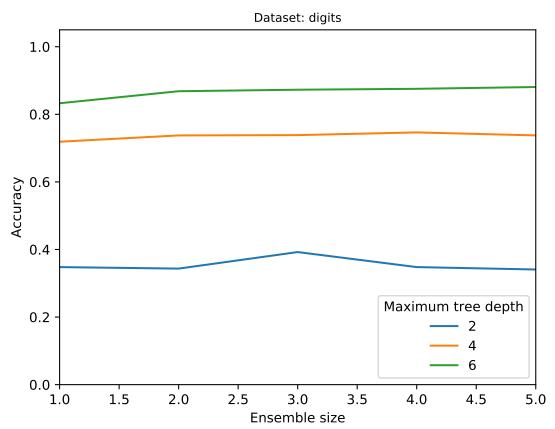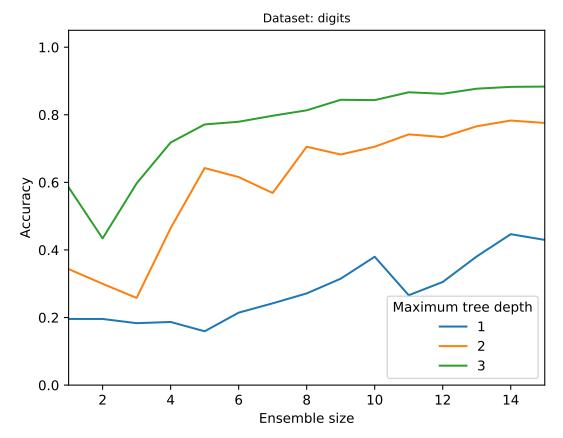
Name: Stewart Kerr

Email: shkerr@wisc.edu

Decision Tree Bagging: Accuracy vs. Number of trees

Dataset: digits

With bagging, the accuracy directly increases with larger maximum tree depth but does not change much with increasing ensemble size. Bagging is useful to combine classifiers with high variance. In the context of decision trees, these are trees that have a large depth - these trees will have small bias (and thus tend to overfit) but will have high variance between different input data. Bootstrap aggregation helps to decrease this variance such that the overall error is low.

Decision Tree AdaBoost: Accuracy vs. Number of trees

Dataset: digits

With Adaboost, we are building successively on previous classifiers. Thus, we can use poor classifiers (that is, decision trees with a small depth - large bias) and still obtain good results by increasing the number of training iterations (ensemble size). From the figure, we've obtained an accuracy over 0.80 using a small tree that has a maximum depth of only three.

For this particular problem, I think that bagging is a stronger method. Increasing the ensemble size increases computation time much more than increasing maximum tree depth, thus bagging (which does not require a large ensemble size) will run much faster that boosting.