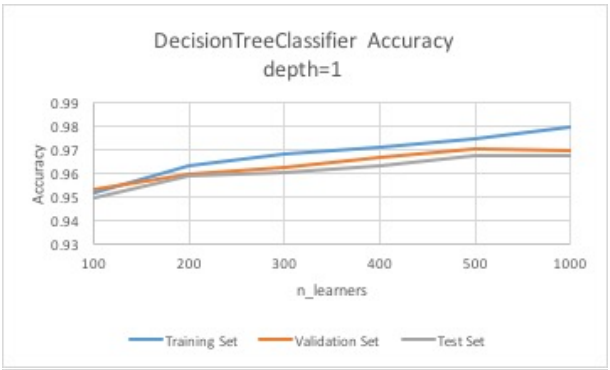
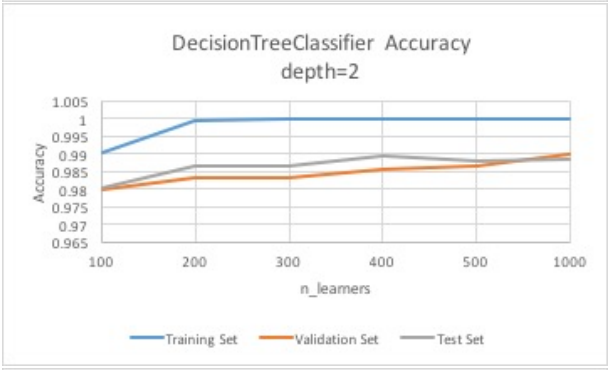
Donovan Guelde
CSCI 5622
Boosting HW

Using Sklearn's Adaboost with the DecisionTreeClassifier, the dataset was analyzed, and the accuracy of distinguishing 4's from 9's was established. At a tree depth of 1, 2, and 3, with n_learners=100, 200, 300, 400, 500, and 1000 for each tree depth, the following results were obtained:

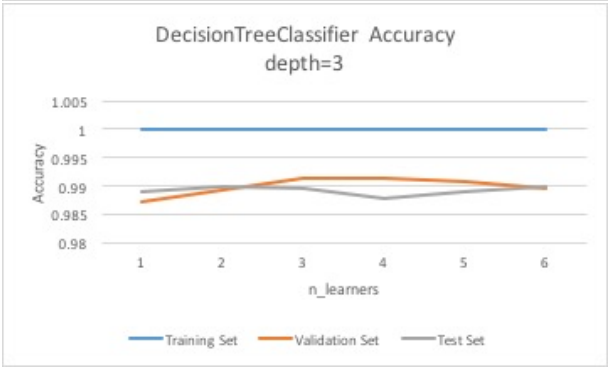
At a tree depth of 1, maximum accuracy was obtained (against a test set) at n_learners = 500 and 1000, with both tests giving an accuracy of 0.9674. There may be some overfitting here, as the accuracy against the training increases as n_learner increases above 500, while accuracy against the test set levels off.



At tree depth = 2, maximum accuracy against the data.x_test set was achieved at n_learners = 400. At this depth, accuracies for all n_learners was within a narrow range, from 0.9804 to 0.9895. There is a definite possibility of overfitting, as accuracy against the training set was 1.0 for all n_learners $\geq$ 200.



At tree depth of 3, all n_learner values resulted in accuracy against the training set of 1.0, indicating overfitting (unless the data is linearly seperable.) Accuracy against the test set appears stable, varying between 0.9879 and 0.9900, with no discernable trend.



The process was repeated, with GradientBoostingClassifier replacing DecisionTreeClassifier in Adaboost. In attempting to find interesting values of n_learners (those showing possible overfitting, etc), I decided to use much lower values of n_learners, from 2^1 to 2^6. n_learner values above this point showed no variation as n_learner continued to increase.

For depth = 1, accuracy against the training, validation, and test set increased steadily from n_learners=2 to n_learners=64, with no obvious signs of overfitting.

At depth=2, accuracy against the training set quickly reached 1.0, at n_learners=8 and above. Accuracy against the validation and test sets, however, continued to increase gradually.// At depth=3, the highest accuracy was achieved, with n_learners = 64, accuracy = 0.9915.