Project 4 Report

Results

```
bagging digits
Training Size (instances): 1437
Testing Size (instances): 360
Decision tree train/test accuracies 1.000/0.794
Bagging train/test accuracies 1.000/0.892
Run Time (s): 9.12986684399948


randforest digits
Training Size (instances): 1437
Testing Size (instances): 360
Random Forest Accuracy 0.908
Run Time (s): 0.06619098399823997


adaboost digits
Training Size (instances): 1437
Testing Size (instances): 360
Decision tree train/test accuracies 1.000/0.794
AdaBoost train/test accuracies 1.000/0.786
Run Time (s): 0.052557677001459524


-----
bagging datasets/mammographic_masses.data.txt
Training Size (instances): 768
Testing Size (instances): 192
Decision tree train/test accuracies 0.833/0.760
Bagging train/test accuracies 0.833/0.760
Run Time (s): 0.48861251099879155


randforest datasets/mammographic_masses.data.txt
Training Size (instances): 768
Testing Size (instances): 192
Random Forest Accuracy 0.766
Run Time (s): 0.022730358999979217


adaboost datasets/mammographic_masses.data.txt
Training Size (instances): 768
Testing Size (instances): 192
Decision tree train/test accuracies 0.833/0.760
AdaBoost train/test accuracies 0.833/0.760
Run Time (s): 0.7428534739956376
```

Analysis

The ensemble approaches show to have really improved the accuracies on some of the runs. The mammographic dataset seems to have been unaffected by the techniques. This is likely due to each attribute having the same weight as the others, none affecting more than the other. This would have the ensemble technique change nothing.