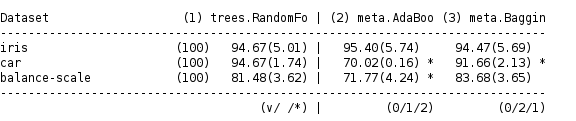
Compsci 361 Assignment 2

Hasnain Cheena  
190411106  
hche737

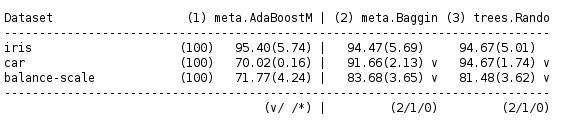
## Part A

3 runs of the experiment were conducted. Within each run the the order the algorithms are run in changes. The results of each run are shown below:

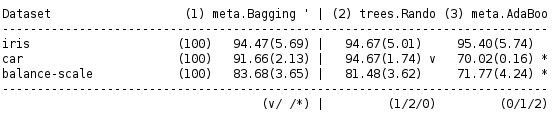
Run 1



Run 2



Run 3



## Part B

**Iris dataset**

We have no evidence (at a 5% significance level) of a difference in performance between the AdaBoost, RandomForest and BaggingTree algorithm on the iris dataset. Therefore, all three algorithms performed similarly well.

This indicates that the iris dataset must have low amount of target label noise and data mislabelling. This is because the AdaBoost algorithm was able to correct true errors and not overfit to errors/noise.

Further, the attributes in the iris dataset are of similar predictive strength. This is because if there was a large discrepency in predictive strength the BaggingTree would have performed poorly. The algorithm considers all the attributes avaliable at each split in the tree meaning the tree’s produced would have been using the same strong predictors at the top and hence be highly correlated.

**Car dataset**

We have evidence (at a 5% significance level) that the RandomForest performed the best, followed by BaggingTree and then AdaBoost.

The AdaBoost algorithm did not perform well hinting there must be significant target label noise and/or data labelling errors in the car dataset. If there is a large amount of target label noise then AdaBoost wil try to classify those more difficult instances with no avail. Secondly, if there is data labelling errors in the underlying data then AdaBoost will incorreclty classify and thus incorrectly weight instances, peforming poorly.

RandomForest performed the best among all three algorithms. This is because the RandomForest’s characteristic to randomly select a subset of the predictors as split candidates at each split in the tree. Thus each tree in the ensemble created by the RandomForest was more different than in the BaggingTree algorithm. This means the tree’s within the RandomForest ensemble were able to correctly classify a relatively bigger pool of cases.

In comparision the BaggingTree algorithm was affected by the underlying discrepency between the strength of predictors in the dataset. The car dataset must contain a few attributes with very strong predictive power (relative to the other attributes). The BaggingTree algorithm considers all the attributes avaliable at each split in the tree, thusthese strong predictors will be used at the top of each tree. Each tree is therefore similar meaning the ensemble is able to correctly classify a smaller pool of cases than the RandomForest ensemble.

**Balance-Scale dataset**

We have evidence (at a 5% signficance level) that AdaBoost performed worse than both RandomForst and BaggingTree algorithms. We have no evidence to differentiate performance between RandomForest and BaggingTree on the Balance-Scale dataset.

The AdaBoost algorithm performed worse than the other two algorithms there must be significant target label noise and/or data labelling errors in the dataset. If there is a large amount of target label noise then AdaBoost will try to classify those more difficult instances with no avail. Secondly, if there is data labelling errors in the underlying data then AdaBoost will incorreclty classify and thus incorrectly weight instances, peforming poorly.

The RandomForest and BaggingTree performed similarly indicating tree’s created by both algorithms were different enough from one another with no/low correlation between them. These algorithms performed well because the bagging mechanism within both created re-sampled training sets that were different enough from one another. Thus the underlying data must have attributes of similar predictive strength.

## Part C

The paired t test results (shown in the tables in Part A) differ depending on which order the algorithms are run in. Each table shows a different view of the results becasue in each run, a a different algorithm has been used as the baseline for the paired t-test. The accuracy of the baseline algorithm is compared to the accuracies of other two algorithms to determine whether there is a significant difference in accuracy and subsequently which algorithm is better. Thus because each run gives you pairwise rankings, three runs in total are needed evaluate the rankings of all three algorithms.

## Part D

I believe the RandomForest algorithm is the more reliable than AdaBoost and the BaggingTree algorithm.

Firstly, RandomForest is more reliable than AdaBoost because of AdaBoost’s underlying boostin mechanism. Boosting is highly sensitive to target label noise/labelling errors. All data sources will contain errors and so this is a huge disadvantage when using AdaBoost.

RandomForest is more reliable than the BaggingTree algorithm. Both BaggingTree and RandomForest perform bagging on the instances but RandomForest additionally performs feature bagging. Thus the tree’s created by the RandomForest algorithm are typically more different than the tree’s created by the BaggingTree algorithm. This means that generally the ensemble created by the RandomForest will be more diverse, covering a wider variety of cases compared to the ensemble created by the BaggingTree algorithm.