PROJECT 3: ASSESS LEARNERS

Christopher Jimenez
cjimenez36@gatech.edu

***Abstract*—** This report will investigate the behavior and performance of four Decision Tree based methods of regression, including bootstrap aggregating methods. I will fit all four methods to the same dataset and evaluate fit as well as the impact of hyperparameters on overall performance.

1 INTRODUCTION

Three experiments were conducted on Istanbul's stock exchange national 100 index data to determine the performance of four Classification and Regression Trees (CARTs) algorithms. The experiments investigated the impact of randomization via random feature selection of splits, as well as bagging on decision tree algorithms, which are prone to overfitting. In my experiments, I evaluated the fit of the models to the data and their resulting accuracy. My findings will show that decision tree algorithms with small leaf sizes tend to overfit and that leveraging bagging can alleviate their tendency to overfit, allowing for better results.

2 METHODS

2.1 Setup

The experiments used the four Decision Tree algorithms described below.

2.1.1 DTLearner

This model is a regression version of JR Quinlan's decision tree model (Quinlan, 1986), using correlation of factors to the target to determine the optimal split factor and then splitting on the median of that factor to maintain balanced trees.

2.1.2 RTLearner

RTLearner is identical to DTLearner, except that rather than using correlation to determine the factor split on, the factor is selected randomly.

2.1.3 BagLearner

The BagLearner was constructed so that it could accept any learner, train many different instances (20 by default) using randomly sampled data and aggregating the output to produce a single prediction.

2.1.4 InsaneLearner

The InsaneLearner was set up to create 20 instances of BagLearner using LinearRegression by default for 20 bags.

2.2 The Experiments

All experiments utilized the same Istanbul stock exchange data, shuffled, with a 60/40 train/test split.

2.2.1 Experiment 1

In order to investigate the potential for overfitting of the Classic Decision Tree learner, I trained and tested 50 instances of DTLearner with leaf sizes ranging from 1 to 50. I then compared and plotted the output for both in and out of sample predictions, with RMSE as my evaluation criteria.

2.2.2 Experiment 2

To understand the impact of bagging on overfitting, I trained and tested 50 instances of BagLearner (with 20 bags) that utilized DTLearner as its weak learner, with leaf sizes ranging from 1 to 50. I then compared and plotted the output for both in and out of sample predictions, with RMSE as my evaluation criteria.

2.2.3 Experiment 3

Finally, in experiment 3, I compared the performance of the Classic Learner (DTLearner) and the Random learner (RTLearner) to explore their strengths and weaknesses. I evaluated them based on Mean Absolute Error and Computation Time (at training). For each metric and model, I trained 50 instances with leaf sizes ranging from 1 - 50.

3 DISCUSSION

3.1 Experiment 1

Figure 1 clearly shows that as the leaf size increases, so does RMSE for both in-sample and out of sample predictions. The low error for in-sample predictions vs the relatively higher error for out of sample predictions where leaf size is 1 suggests that very small leaf sizes do not generalize well. This is because for small leaf sizes, the tree becomes quite deep and more specific, hindering its ability to generalize. As the leaf size increases, the RMSE of in-sample and out of sample predictions begin to converge (and intersect at leaf size = 40), indicating less or no overfitting.

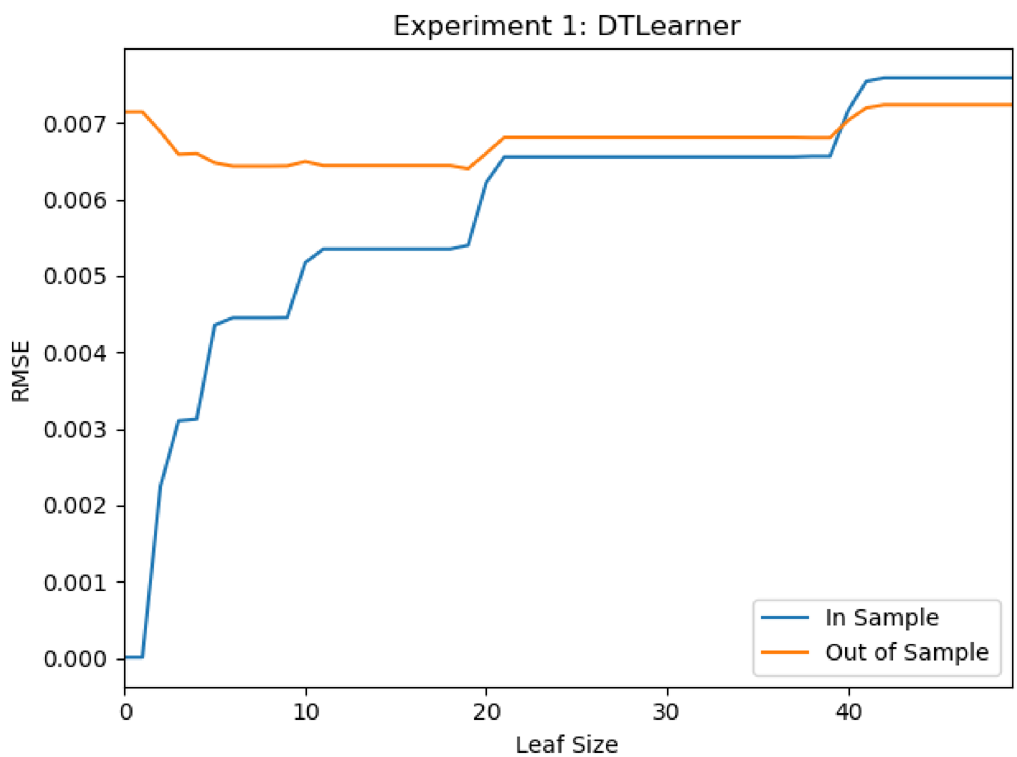

Figure 1—RMSE of the in-sample and out of sample predictions of the DTLearner model with leaf sizes ranging from 1 - 50.

3.2 Experiment 2

Introducing bagging to the DTLearner resulted in less overfitting for smaller leaf sizes. The randomness introduced by bagging increased the model's ability to generalize for out of sample predictions. We are trading off bias for variance. Overfitting still occurred at small leaf sizes, such as leaf size = 1. As seen in Figure 2, the initial gap of RMSE between in and out of sample predictions was smaller and converged sooner when compared to Figure 1. While bagging may reduce overfitting, it is unlikely to completely eliminate it. For example, for small n-counts of sample data are more likely to overfit because they are less likely to capture the full variance found in the population. A biased sample cannot be corrected for via bagging alone.

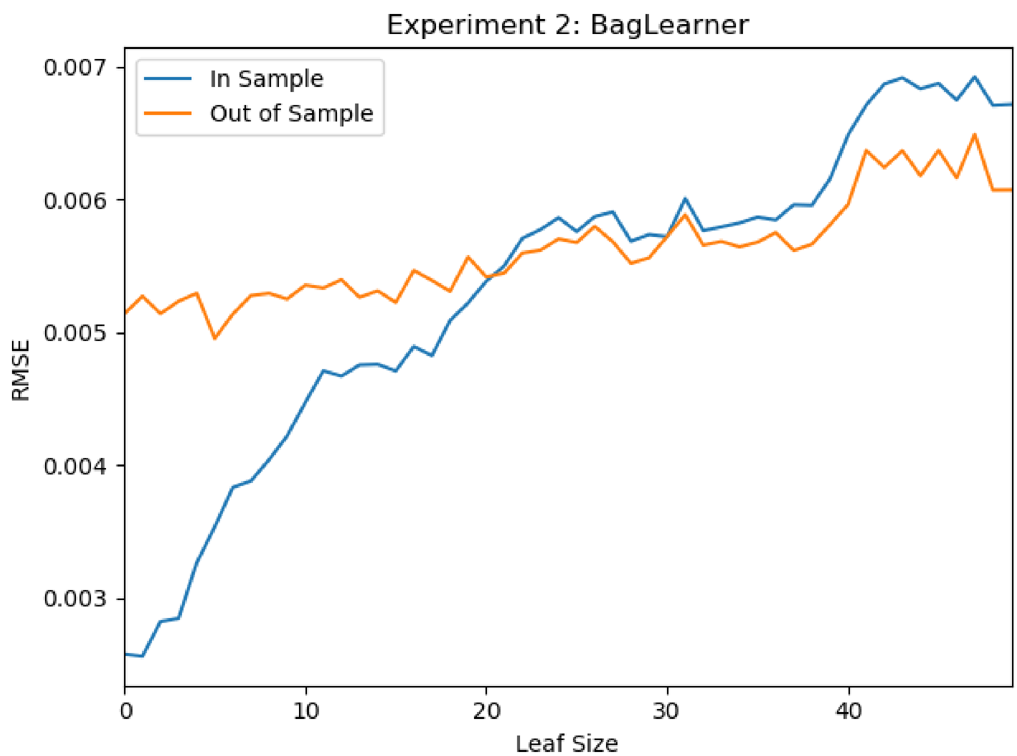

Figure 2—RMSE of the in-sample and out of sample predictions of the BagLearner model using a classic Decision Tree with leaf sizes ranging from 1 - 50.

3.3 Experiment 3

Classic Decision Tree learners and Random Decision Tree learners each have their strengths and weaknesses. When compared directly, Classic learners were generally more accurate than Random learners as seen in Figure 3. In this case, this is due to the classic learner having a specific criteria that ensures the splits are based on some measure of predictive power (correlation), whereas the

Random learner was completely random. The power of the Random learner comes through when combined with bagging because of the Central Limit Theorem. With a high number of random samples, the variance will stabilize and produce good results with low bias. For this reason, without bagging, the Classic learner is likely to be better than the Random learner. However, with bagging, the Random learner is likely to be more accurate, as the Classic learner would potentially have more bias due to the criteria used to split the trees.

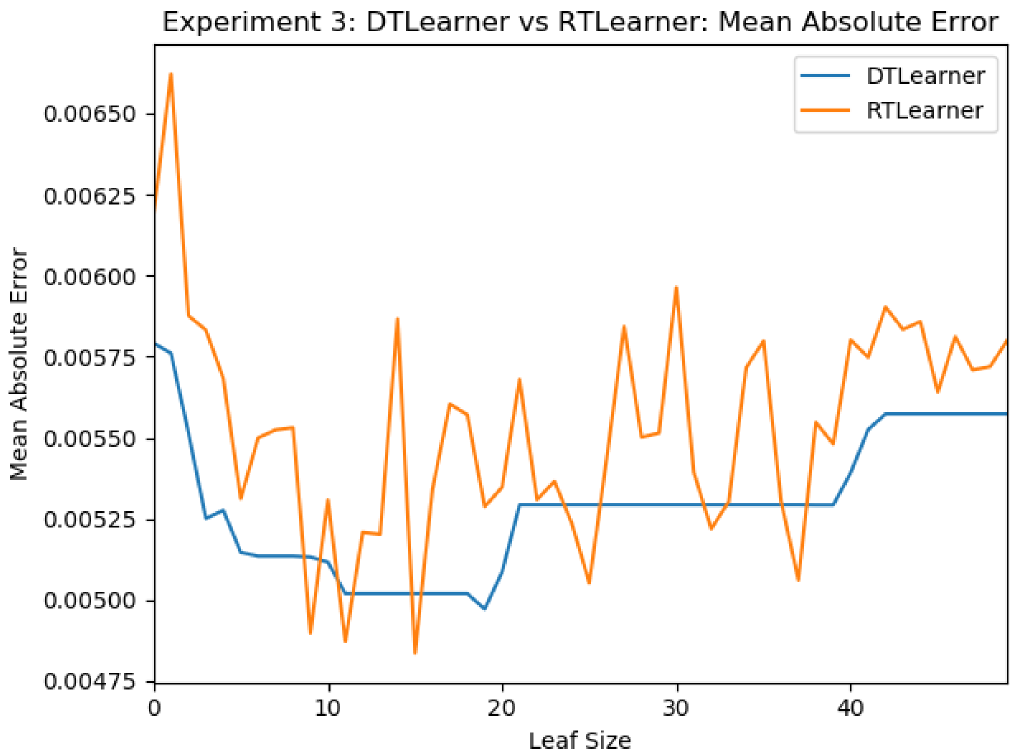

Figure 3—MAE of out of sample predictions of the Classic and Random learners with leaf sizes ranging from 1 - 50.

Within this experiment, I also evaluated the two models based on computation time. In this respect, the Random learner was the clear winner across all leaf sizes, although the difference shrunk as the leaf size increased as seen in Figure 4. This is due to the additional, more complex calculations required by the Classic learner to determine splits. The smaller the leaf size, the deeper the tree, which results in more calculations of split criteria, so the Random learner is faster because its split method is comparatively simple. In this respect, the Random learner will always be faster, although the difference is small for large leaves.

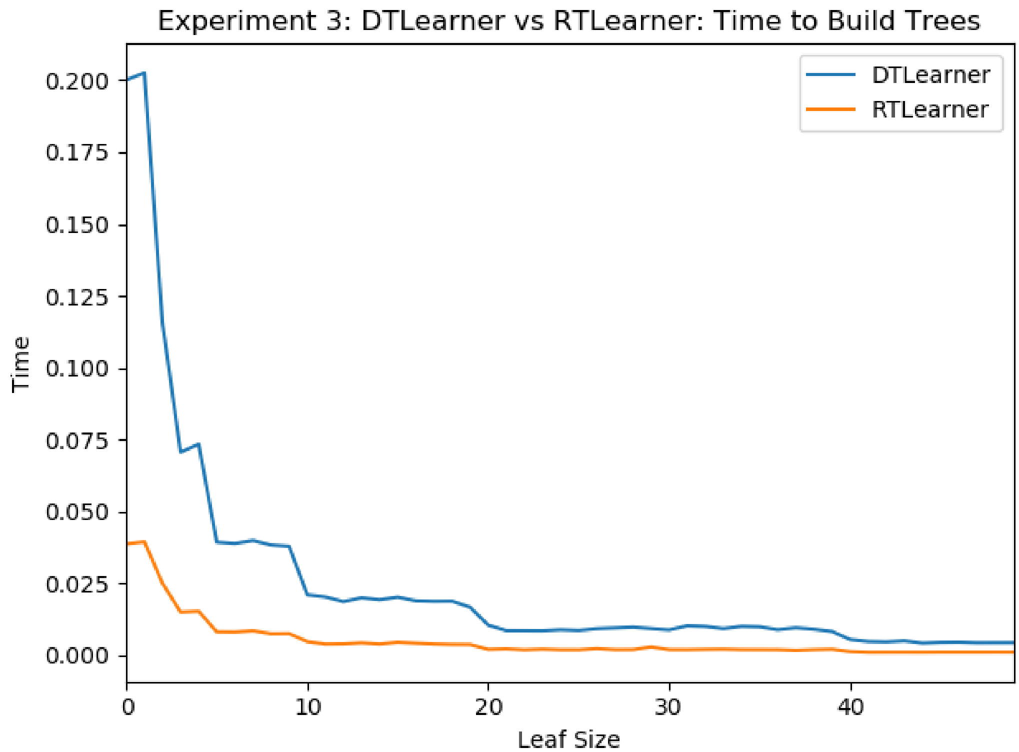

Figure 4—Time to build the Classic and Random learners with leaf sizes ranging from 1 - 50.

3 SUMMARY

Throughout these experiments, it was clear that small leaf sizes results in deeper trees, which tend to lead to overfitting. Overfitting can be mitigated by randomness, but the randomness is most beneficial in large samples, like seen in bagging. The larger sample sizes from bagging are needed for the variance to properly stabilize and produce good results. As seen with the RTLearner, a single randomized tree is inconsistent in its predictive power, regardless of leaf size. Although they are more computationally expensive due to their iterative nature, learners that utilize bagging outperformed other models in terms of fit and accuracy and are the recommended approach over single decision tree learners that are vulnerable to overfitting.

4 REFERENCES

1. J.R. QUINLAN (1986), Induction of Decision Trees, Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia