Project 3 Report:
Assess Learners

Youssef Sultan
ysultan@gatech.edu

***Abstract*—**In this report, a series of experiments are executed in order to understand the discrepancies between different regression decision tree models. Each model utilizes a different methodology in its training and data sampling approach which serves to create predictions for input data minimizing error in different ways.

Introduction

Decision tree models are widely used in machine learning approaches to make predictions on unseen data using previous data. In this paper, a comparative performance review of decision trees, random trees and bagging trees are executed in order to understand the caveats of each with respect to increases in leaf size. The hypothesis that can be instantiated is that for any type of decision tree, regardless of the sampling methodology as leaf size increases the variance will also increase.

Methods

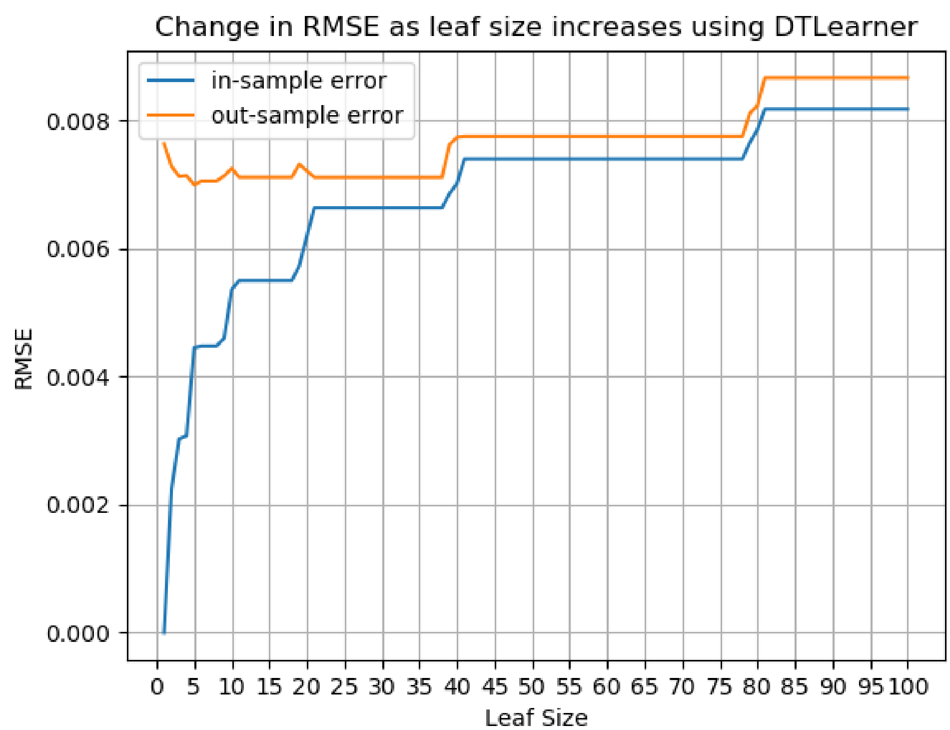
In the first experiment, a single decision tree is trained on the Istanbul dataset and evaluated using RMSE as the metric for 100 iterations, where in each iteration the leaf size increases by 1. The in-sample and out-sample errors for each change in leaf size are depicted linearly on a graph to make comparisons from 0 to n where n is the number of total iterations of the experiment.

In the second experiment, a bagged learner decision tree is trained on the Istanbul dataset and evaluated using RMSE as the metric for 100 iterations, where in each iteration the leaf size increases by 1. The in-sample and out-sample error for each change in leaf size is depicted linearly on a graph to make comparisons from 0 to n where n is the number of total iterations of the experiment. The amount of bags used for this experiment equates to 20.

In the third experiment, R^2 and MAE are used as evaluation metrics to compare how effective a single decision tree and a single random tree are amongst each other. The two evaluation metrics selected are used to provide a comparative overview of how one could be trying to solve different business questions and whether the same models trained on the same data could answer different questions. In this experiment, both models are trained on the Istanbul dataset for 100 iterations, where in each iteration the leaf size increases by 1. Only out-sample error is plotted to show a more robust comparison.

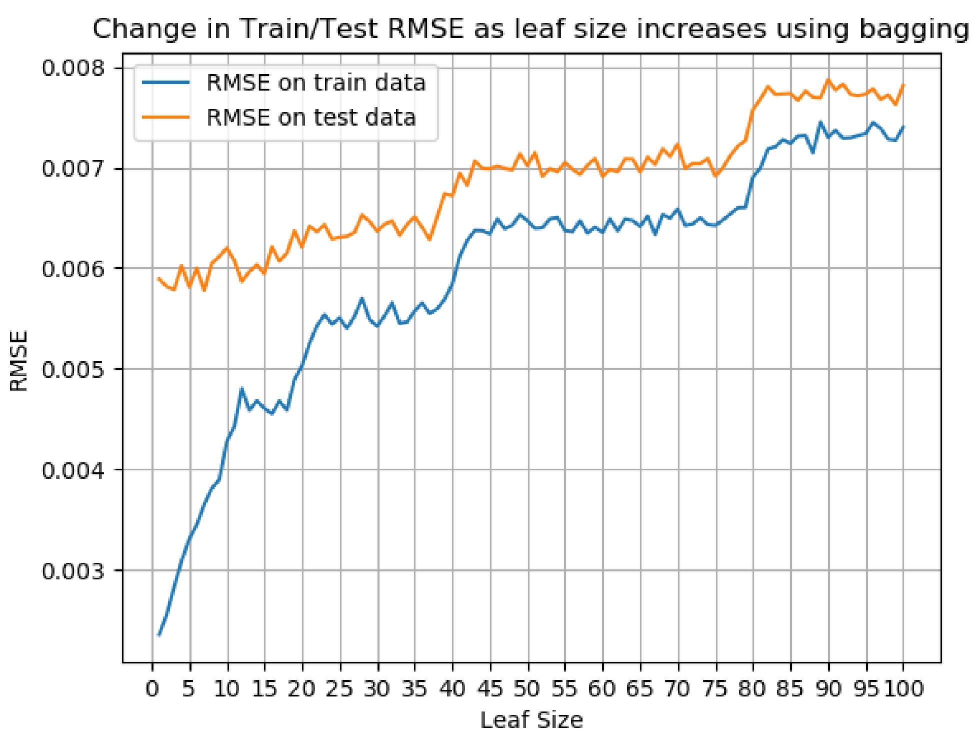
Question 1

When testing a single decision tree model using RMSE as the evaluation metric on the Istanbul dataset, overfitting can be seen in this experiment where the leaf size is at 20 and both lines start to diverge from each other. Where the leaf size is 20 or greater, the predictions do not generalize as well as they did at smaller leaf sizes due to high variance in the model for the out-sample data. The direction of overfitting is positive and can be seen where the RMSE is greater than 0.006, as the trend follows higher than the minima initially found where the leaf size is 19.



Question 2

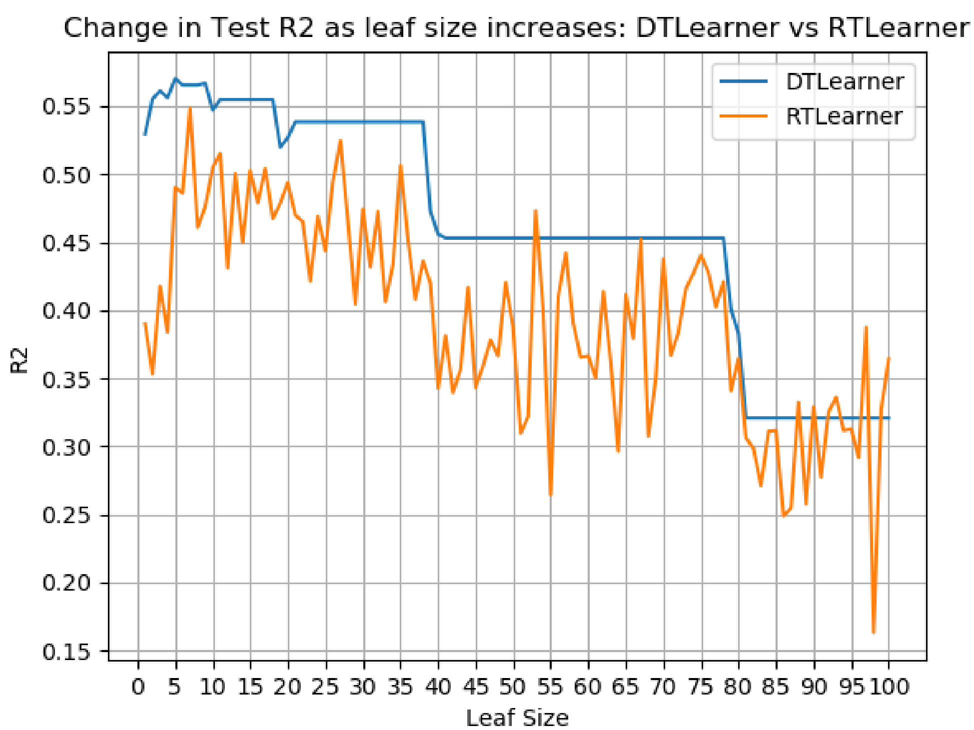
Bagging can reduce overfitting and allow for better generalization of the input data as evidenced by the reduction in RMSE in the Istanbul dataset of the out-sample data compared to the RMSE of a single decision tree. Where the leaf size is 20 the RMSE is below 0.005 which is less than the RMSE of the single decision tree with the same leaf size of 0.006. The reduction in error occurs because the model takes a different random sample with replacement of the data for each bag; averaging out the predictions of all bags contains more information than a single decision tree. This decreases the variance of the model, however as the leaf size increases the error will increase as evidenced by the positive slope of the out-sample RMSE. The positive correlation between leaf size and RMSE confirms that bagging can not completely eliminate overfitting with respect to leaf_size because regardless of how the model is trained, if the leaf size is too high the model won't generalize well to incoming data due to high variance. This also shows the benefits of pruning a model in hopes to reduce the error, if this were to be looked at from a hyperparameter tuning standpoint, in this case pruning the tree at a leaf size of 7 would give the most optimal results.



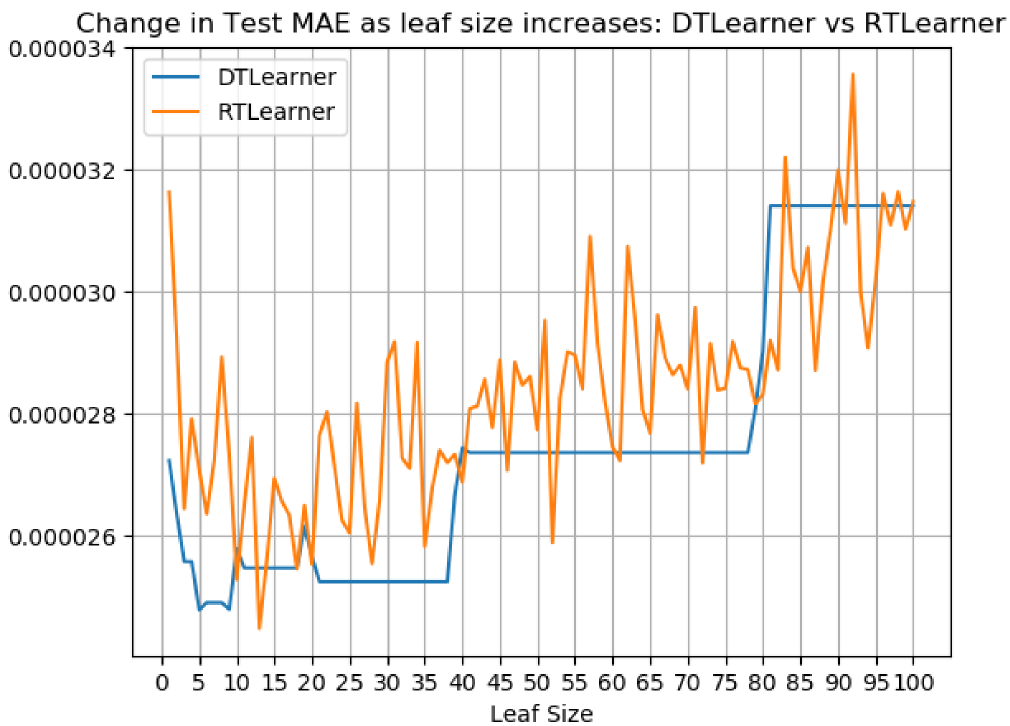
Question 3

In this experiment, the decision tree and random tree are trained on the Istanbul dataset and evaluated using RMSE and MAE for 100 iterations, where in each iteration the leaf size increases by 1. The out-sample error is shown for each leaf size of each model to make comparisons.

The evaluation metric R^2 is used on the Istanbul dataset to compare whether a single decision tree is better than a single random tree by showing how the prediction output fits the target output. The higher the R^2, the smaller the difference is between the fitted values and the observed data. It can be seen that the single decision tree learner is better than the single random tree learner when it comes to fitting the data as evidenced by having the highest R^2 overall, specifically where leaf size is 8. This makes sense because the single decision tree picks a feature based on correlation. Since it already scans for correlation before it finds a value to split on, it will fit the observed data values better than the single random tree. This can be observed with this metric especially given that it is based on correlation.



Additionally, in this experiment, the evaluation metric MAE is used to calculate the magnitude of an error on average. The lower the MAE the more the predicted values match the actual target values. In this case, it can be seen that the single decision tree is better than the single random tree as it has the lowest MAE for more iterations than the random tree learner. Since the random tree finds the best feature to split on randomly, each trial will have a different variation in how the model is trained. It can be seen in trials where the leaf size is from 0 to 40, the decision tree learner starts to overfit, while the random tree learner also shows a positive trend towards higher error. This makes sense since the single decision tree has a static nature in its training approach in that it always splits based on the feature with the highest correlation.



There is not a one-size fits all learner for all use cases when it comes to making predictions since the in-flow of data can be different in its structure, linearity, or stochasticity. Depending on the use case at hand, if the input data is non-linear by nature some models that have more randomness in their implementation may be a better fit, while if the input data is more linear some basic models may perform better than complex ones allowing for more interpretability. Although the comparison of models was tested on the same data, both models won in their own ways when comparing different evaluation metrics and this shows not only why it is important to pick a proper evaluation metric but also how each model selected depends on the question that is trying to be solved at hand.