Project 3 Report
Machine Learning for Trading
Jie Lyu
903329676

Question 1

Yes, overfitting occur with respect to leaf_size. For Istanbul.csv, overfitting starts when leaf_size < 5. From Figure 1, we can see that the RMSE decreases as leaf size increases from 1 to 5, indicating an overfitting before 5. After 10, the the RMSE increases, indicating an undercutting. Around 5 to 10 lies the global minimum, which indicates the optimal fitting size. Plots here are generated from Istanbul.csv.

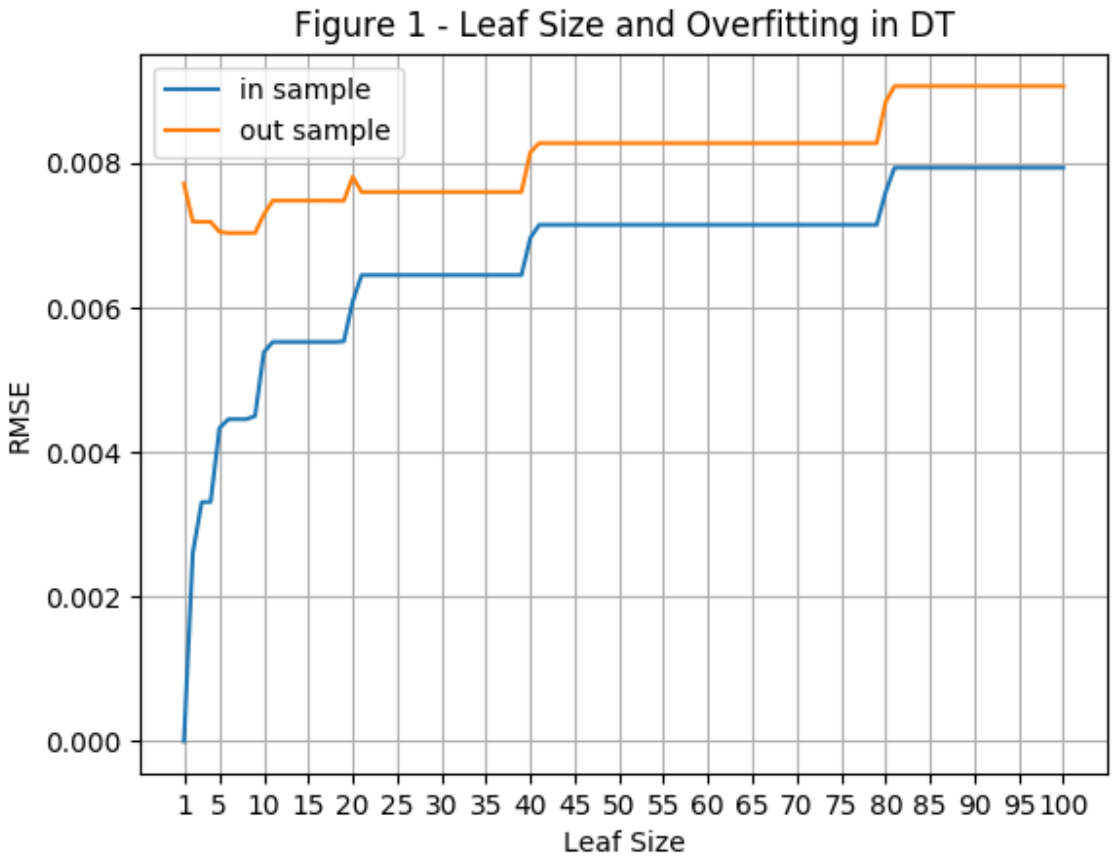
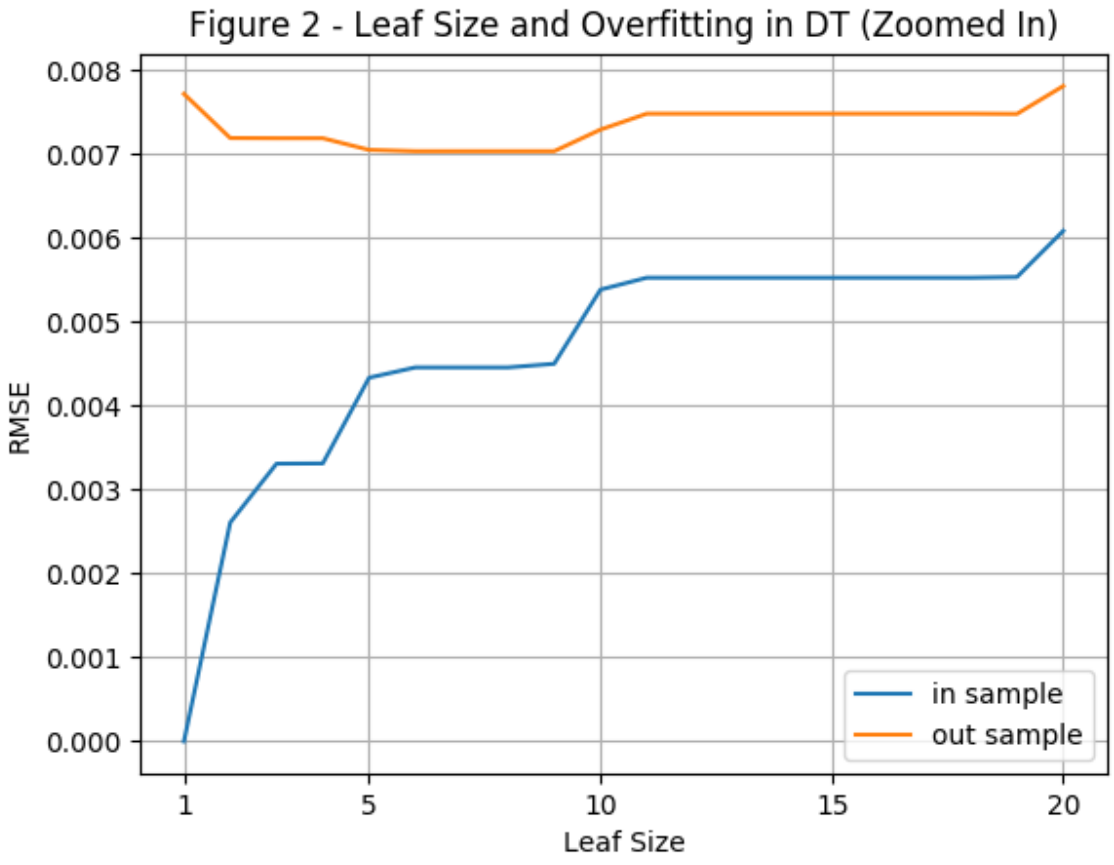
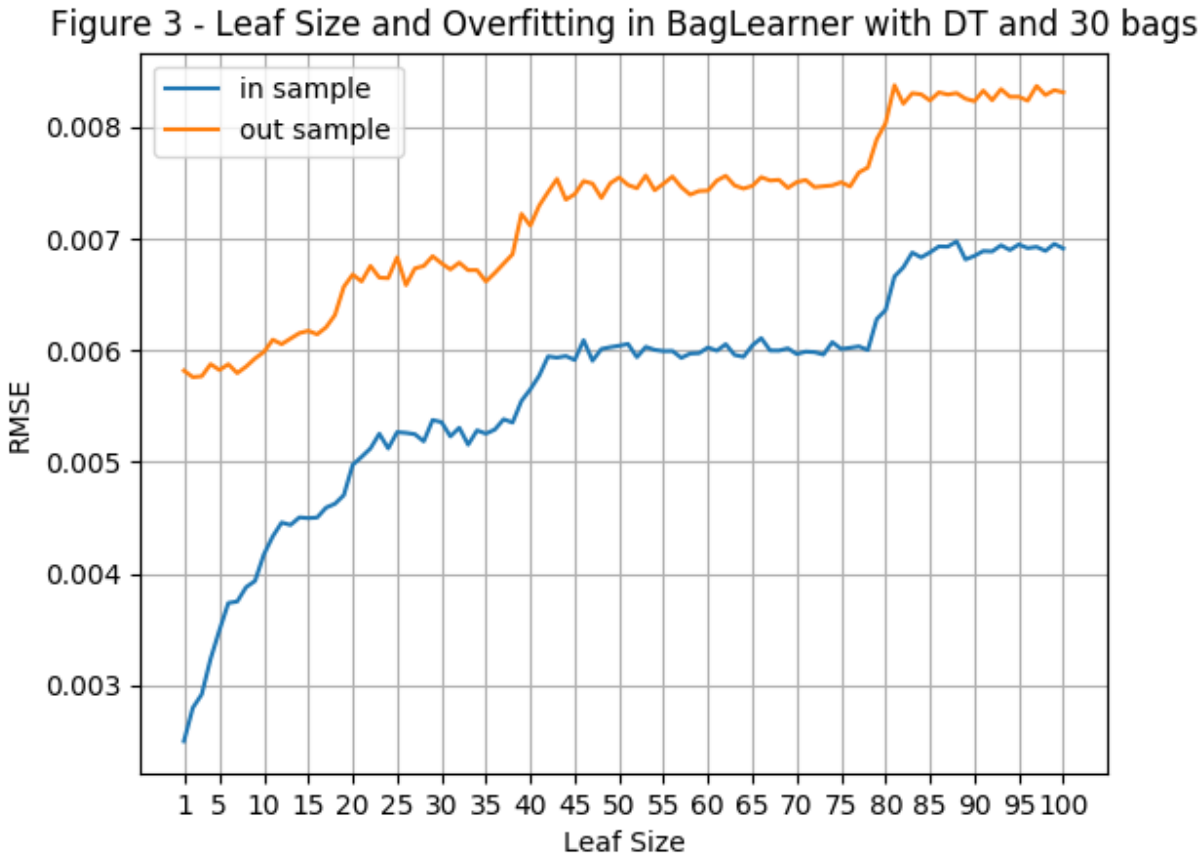


Figure 2 here gives you the zoomed in view of the exact Figure 1 with leaf size from 1 to 20. In general, smaller the leaf size, more descriptive the decision tree model. With a leaf size of 1, the tree generated from training data is almost perfectly descriptive of the training data with an almost zero RMSE for in sample data. However, a perfectly descriptive model is not ideal for prediction because a good model should work well on unseen data. The tree generated with such a small leaf size is not generalized enough for predicting new data outside the training data. Overfitting happens here. As the Figure 2 illustrated, the out sample RMSE is not the lowest at leaf size 1. It gets better when we increase the leaf size to around 5, and gets worse again when leaf size is bigger than 10, where under-fitting happens.



Question 2

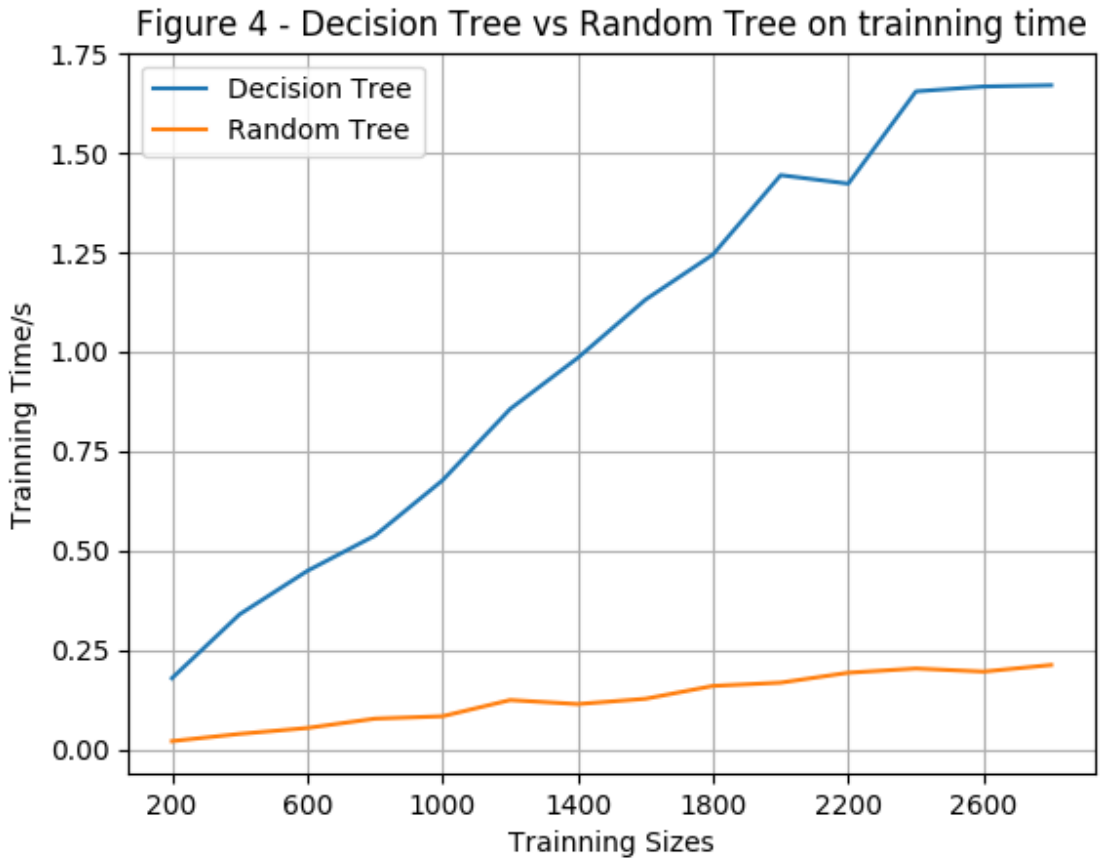
Yes. Bagging eliminates overfitting. Figure 3 is generated from BagLearner with a bag size of 30. As the out sample line illustrated, the minimal RMSE for out sample happens when leaf size is 1. As Figure 1 and Figure 2 shows, overfitting happens at leaf size 1 without bagging. Now, leaf size 1 becomes the optimal leaf size that leads to a global minimal RMSE. So, we can say that bagging eliminates overfitting. Plots here are generated from Istanbul.csv.



Question 3

First metrics we tested out is the training time. Speed is a very important factor when choosing a model because training a large dataset with an inefficient algorithm can cost unnecessary money and time. In this experiment, various size of training sets are passed in decision tree learner and random tree learner, and the training time is plotted. winequality-white.csv is used here and for Figure 7 since it contains more data entries.

From the graph, we can see that for both learner, training time is proportional to the size of training data. However, the random tree learner is significantly faster. At a training size of 1400 entries, DT learner takes 1.0044s and RT learner takes 0.1215s, making RT learner 7.3x faster than DT learner. This is due to the fact that DT learner picks the best feature each time but RT just choose a feature randomly.

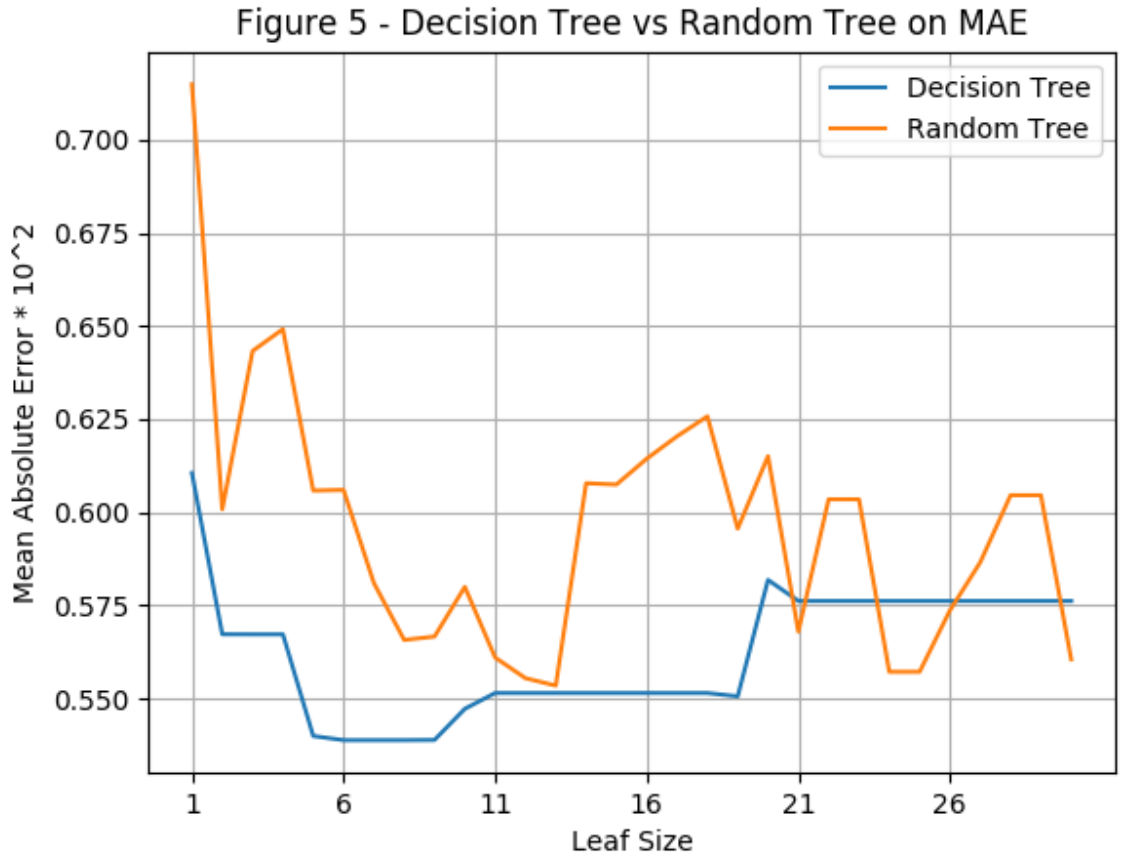


Secondly, we compare the accuracy of the two trees. Mean absolute error is chosen here as a accuracy measure. Unlike RMSE, there is no squaring action here, so big error is not penalized that much. The formula is listed below. Plots here are generated from Istanbul.csv.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x|$$

In Figure 5, MAE of out sample is plotted against leaf size. Only out sample is tested since it is a better indicator for real world performance. In general, DT learner performed better than RT learner as most part of DT learner line is below the RT learner line. At leaf size 1, the MAE is 0.6105 for DT learner and 0.7149 for RT learner, meaning DT learner performs 17% better than RT learner.

The performances of DT learner is similar to its performance in Figure1 and Figure 2. Only one local minima shown around leaf size 5 to 10. The line is smooth and clearly directed. On the other hand, RT learner shows more randomness. The line is stochastic with 5 local minima.



In conclusion, one single DT learner performs better than one single RT learner but it is costly to train. However, we rarely use a RT leaner by itself. When bagged together, its randomness will be decreased. As shown in Figure 6 and Figure 7, a 7-bagged RT takes less time to train and performs better than a DT.

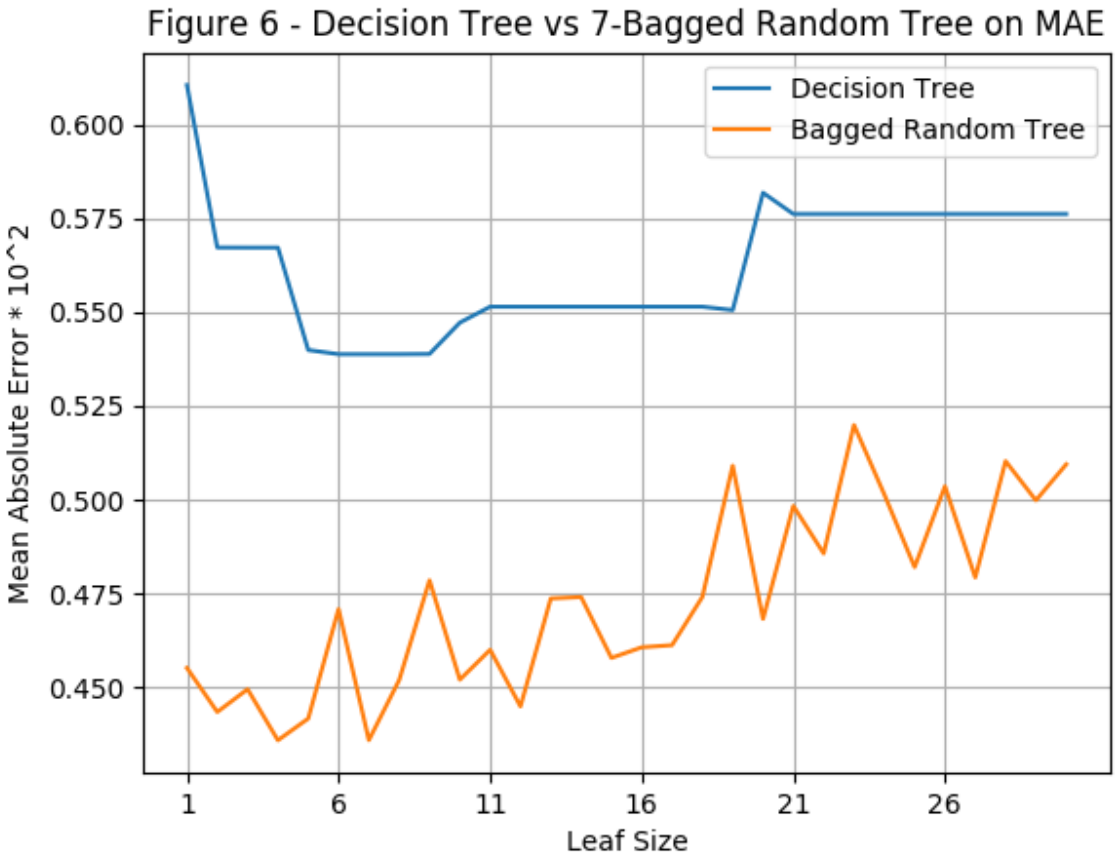


Figure 7 - Decision Tree vs 7-Bagged Random Tree on trainning time