

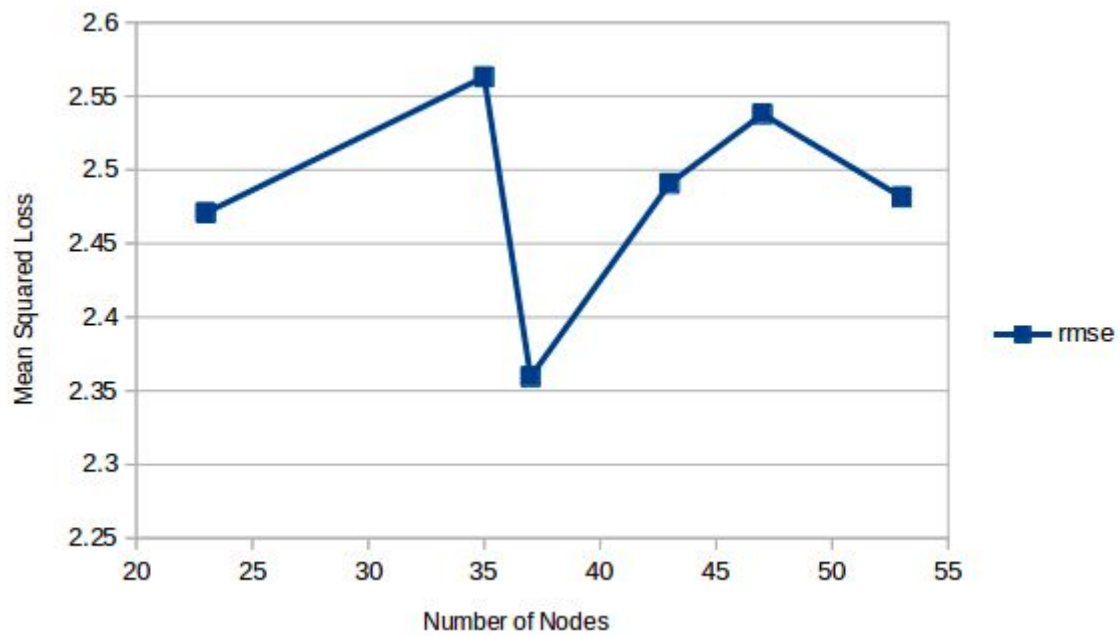
Loss Values as seen on Kaggle:

Competition 1 - 2.98226

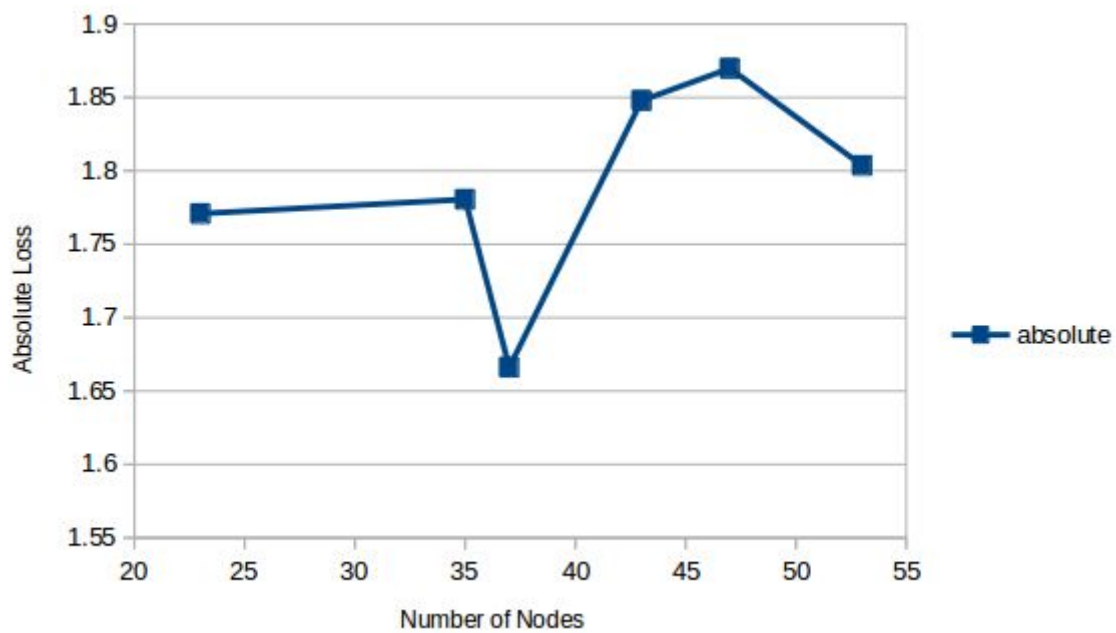
Competition 2 - 0.72603

Graphs-

1) Mean Squared Loss



2) Absolute Loss



Times

1)Kaggle 1 (competition 1)

Training time - 1.095 s

Test time - 0.372 s

2)Kaggle 2 (competition 2)

Training time - 19.2273 s

Test time - 2.05 s

Implementation Details

- Since the output variables to be predicted were continuous variables we implement a regression tree.
- In decision trees used for class predictions, we used entropy or gini for evaluating information gain. However, for evaluating gain for continuous output variables we used variance which mapped how much each element deviated from the predicted value.
- The stopping criteria used for creating leaves is the value of min_leaf_size.
- The regression tree is implemented using the Sprint algorithm as discussed in class.
- The tree is then pruned to prevent overfitting. Various train-validation splits were experimented with. We settled on a split of 5:1.
- For reporting errors for various numbers of nodes, we used a train-validation-test split of 4:1:1