Report

Devanshu Singh Gaharwar-16D070042
Abhiraj Kanse - 16D070006

August 2018

1 Part-A

The best loss value obtained for the smaller dataset given is 2.66710 obtained from kaggle.

2 Part-B

1-fold cross validation is used for pruning. The graph below shows the error vs number of nodes for both loss functions.

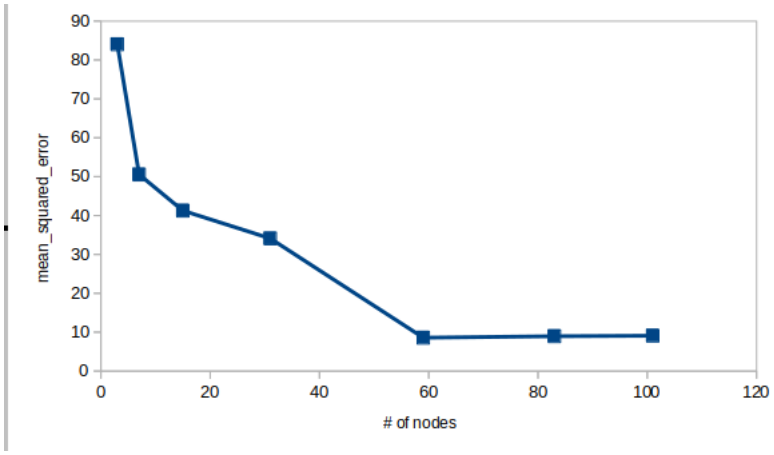

Figure 1: Graph for Mean Squared Loss

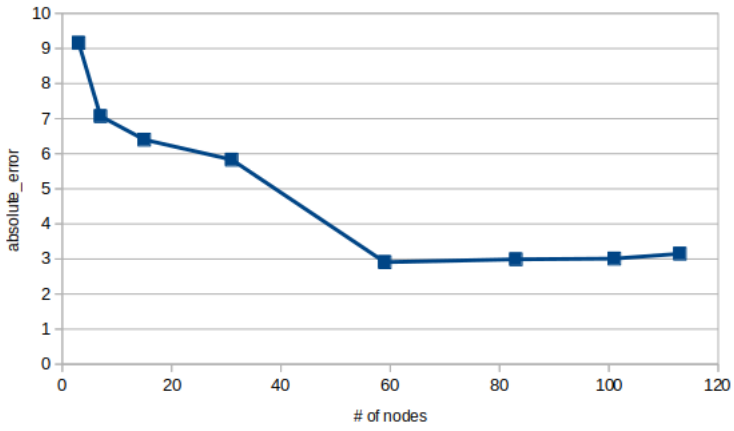

Figure 2: Graph for Absolute Error

3 Part-C

The best loss value obtained for the bigger dataset given is 0.80426 obtained from kaggle.

4 Part-D

Traning time for smaller dataset= 1.303 seconds
Inference time for smaller dataset= 0.003 seconds
Traning time for bigger dataset= 90.714 seconds
Inference time for bigger dataset= 0.009 seconds

The parameter value of min_leaf_size used is 1 to get the minimum error.

Implementation - To implement this Regression tree we first coded the basic functions required for computation and then built the tree recursively using these functions.
To check for the leaf node we conditioned on the given parameter min_leaf_size For Pruning we defined a parameter max_depth in our build_tree function which defines the depth of the tree. Now we can optimize our performance using this parameter by checking error in 1/3rd data given for training.