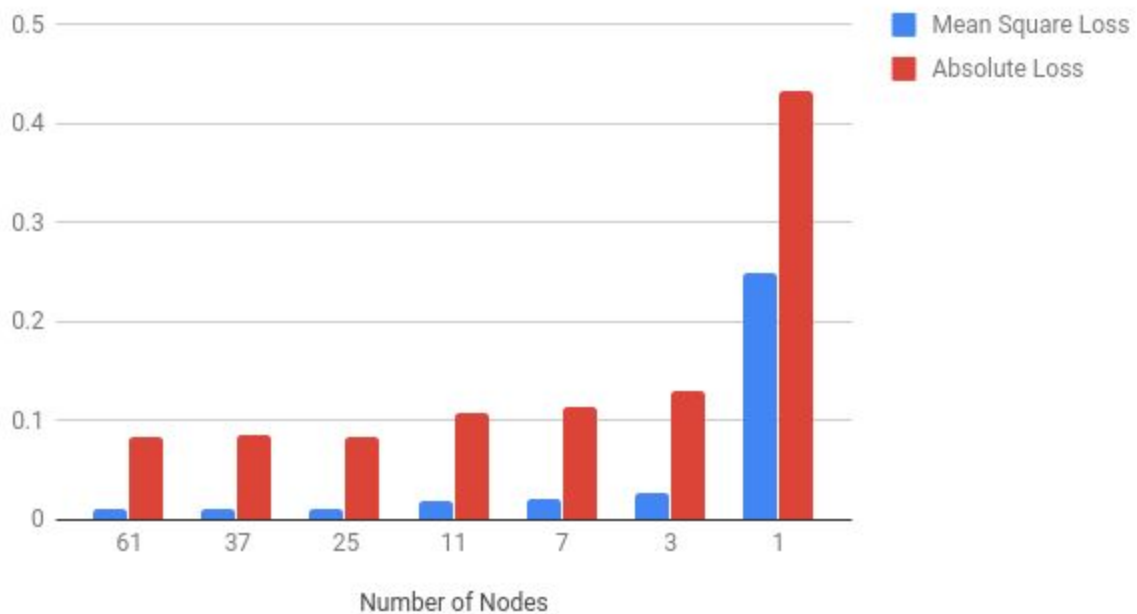


Assignment 1 Report
16D070011 - 16D070013

1.) Loss vs Number of nodes:

Number of Nodes	Mean Square Loss	Absolute Loss
61	0.011	0.084
37	0.0114804	0.0854265
25	0.011235	0.0841232
11	0.0199515	0.108491
7	0.02208	0.114509
3	0.0273569	0.13037
1	0.248878	0.432973

Mean Square Loss and Absolute Loss



2.) Best loss values (mean square error):

Dataset1 : 7.9

Dataset2 : 0.900072

3.) Training and Inference time

Dataset1:

Training time: 92.8 ms

Testing time: 0.45 ms

Dataset2:

Training time: 2.985 s

Testing time: 21.021 ms

4.) The regression tree was implemented in C++ with each node of the tree represented/implemented by an object of the class "Node". Thus, each Node carries with it the attribute used for best split and two pointers to its left and right subtrees respectively (Node* lchild and rchild). Apart from this it also has functions to find the best split and create further subtrees according to it. Pruning was done by a method as discussed in class. Starting from the leaves, if the weighted mean square error of the parent is less than that of both its children then the children are pruned and the parent is made a leaf and so on. The prediction of a leaf node is equal to the average of the data values belonging to that leaf. For controlling the number of nodes the parameter used was the minimum number of data entry points/instances required to be present in each node. Thus, if this parameter is increased, the number of possible splits decreases, hence number of nodes decreases.