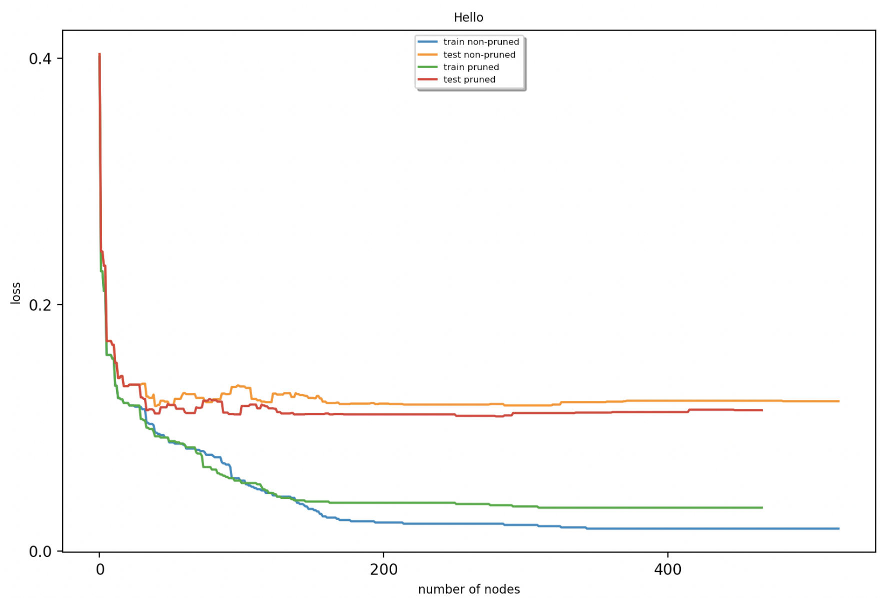
Project Report 6

1. The loss for non-pruned trees on training data sets is better than the loss for non-pruned trees on training data because non-pruned trees are specifically made to minimize the empirical risk on training data set. The loss for pruned trees on training data sets is worse than the loss for pruned trees on training data because pruned-trees are the result of both training and validation dataset so that the tree is generalizable to other data sets.



Above is the result with the maximum depth 40. Comparing test results for pruned and non-pruned, we can confidently conclude that our pruning was effective. as it achieved lower loss. In addition, comparing the loss for three loss calculation techniques, I noticed that node score entropy tends to return the lowest loss among the three techniques.

2. I observed that the curve with all 1-15 node depth looks like a flipped version of log functions in shape. That said, as we increase the max depth more, we reduce less loss. On the other hand, adding the first few depths has a great impact on reducing the loss. This trend is likely to be the result of the fact that we implemented a greedy algorithm to keep finding loss minimizing attributes of the data and so the earlier nodes are split by indices that have greater impact on reducing the loss. However, since it is greedy, we also observed some fluctuations in the trend.
3. I would totally agree with the author's idea because representation, loss, and optimization are the core of the machine learning process and everything is dependent on the metrics that humans give to the machines. So in terms of the importance, yes, I think that Goodhart's Law grows increasingly relevant to ML. However, at the same time, I am wondering as to how we would create an algorithm that would create an algorithm to learn to produce answers that humans find appealing as opposed to accurate answers as we also have to give metrics for that algorithm as well.