YOUR ID HERE! **Homework Report 6** CS1420 – Spring 2019
March 8, 2023

Problem 1

Comment on the results of your final program. Discuss the differences in training and test error of pruned and non-pruned trees. Which measure of gain most effectively reduced training error? Was pruning effective? Use tables or graphs to demonstrate your findings

Overall, the training error is lower than the testing error, and the error of pruned tree is lowered than that of the non-pruned tree. We can notice that the effect of the entropy gain function and the gini function is much better than error, with gini a little better than entropy. This is because the loss decreases at a faster rate to a lower result, and in the spam dataset, specifically, the entropy and gini have more effective prevention of overfitting. We can therefore conclude that pruning is effective, since the loss of the pruned tree is overall lower than that of the unpruned tree.

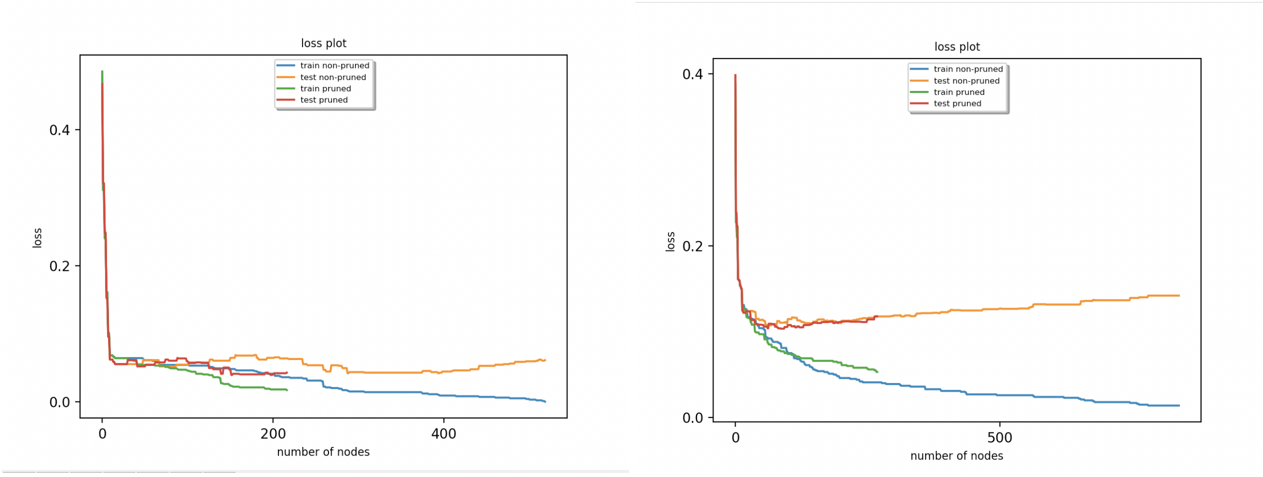

Figure 1: Chess dataset and Spam dataset, error

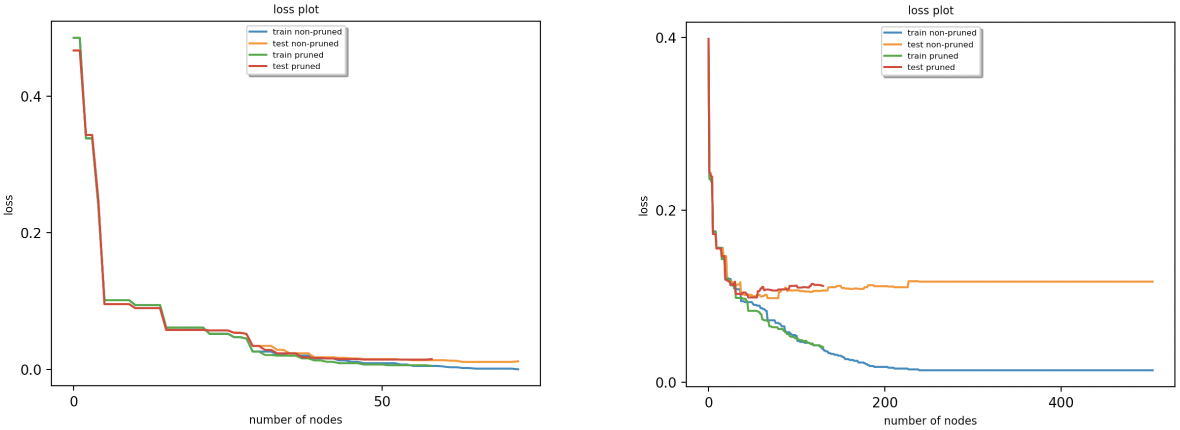

Figure 2: Chess dataset and Spam dataset, entropy

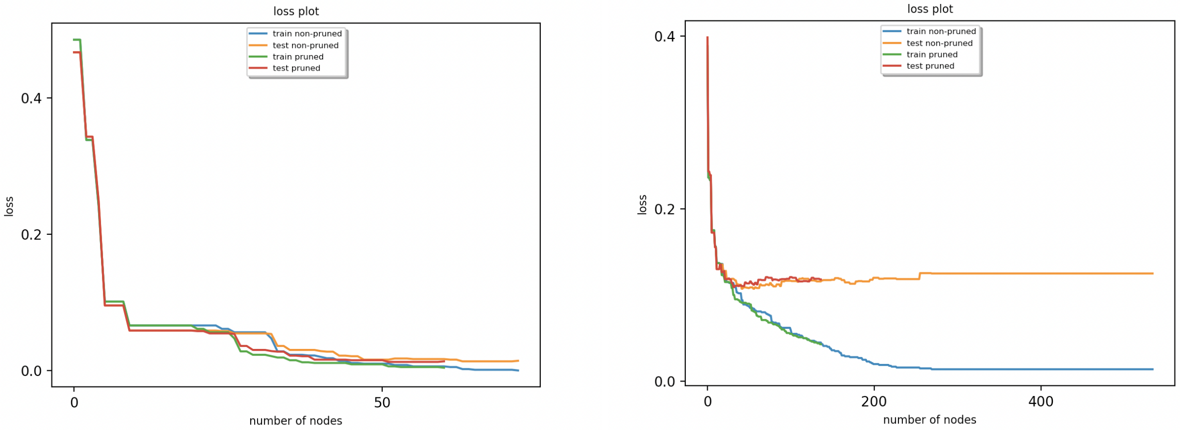

Figure 3: Chess dataset and spam dataset, gini

Problem 2

Using the spam.csv dataset, plot the loss of your decision tree on the training set for trees with maximum depth set to each value between 1 to 15. For these plots, the trees should not be pruned and you can use the entropy gain function. Discuss any trends you find and attempt to explain them in three to five sentences.

As the maximum depth of the tree increases, the model becomes more complex and able to fit the training data more closely. Therefore it is clear that from the plot, the loss decreases at a steady rate from higher to lower with the maximum depth increasing. We can also notice that the decrease rate also decreases as the maximum depth grows, in other words, the reduction of loss decreases, so the effect of increasing maximum depth to reduce loss may reduce as maximum depth goes higher.

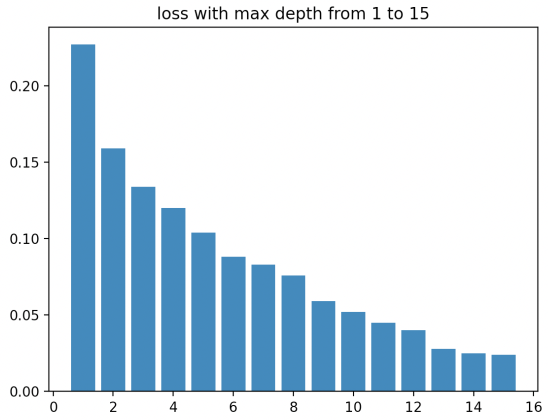

Figure 4: Loss of trees with maximum depth set between 1 and 15

Problem 3

What are your initial thoughts on reading this passage? Do you agree with the author? Why or why not? Please elaborate your answer.

My feelings upon reading this passage are that it opens a new perspective for thinking about the potential risks in ML algorithms, since I have never considered that AI might learn to adapt to human preferences. Goodhart's Law states that when a measure becomes a target, it loses its effectiveness as a measure because people begin to optimize for it to the detriment of other important factors. In the context of machine learning, this could happen when the optimization of a metric becomes the primary goal, rather than the accuracy or effectiveness of the model in solving the problem at hand. In such a situation, the model may become overfit to the training data, leading to poor generalization and potentially harmful outcomes when deployed in the real world. Moreover, the part that states that AI may produce answers that humans find appealing is also theoretically possible. If the metrics used to evaluate the performance of the model do not accurately capture the desired outcome, the algorithm may choose to fit to human preference instead.