

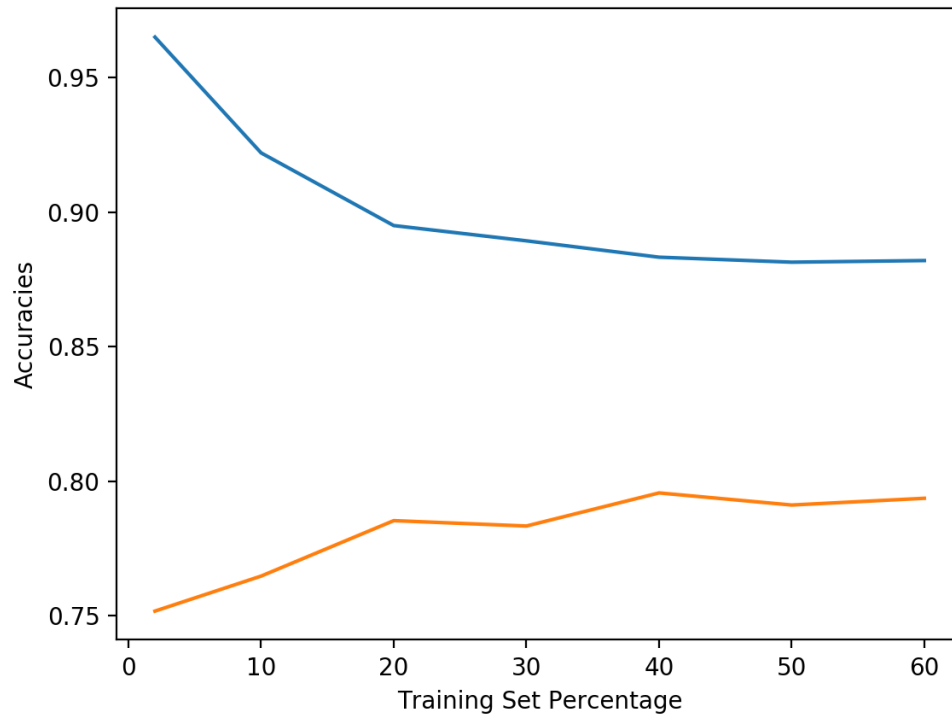
CS 373 Homework 3 (Using 2 late day)

By: Siddharth Shah

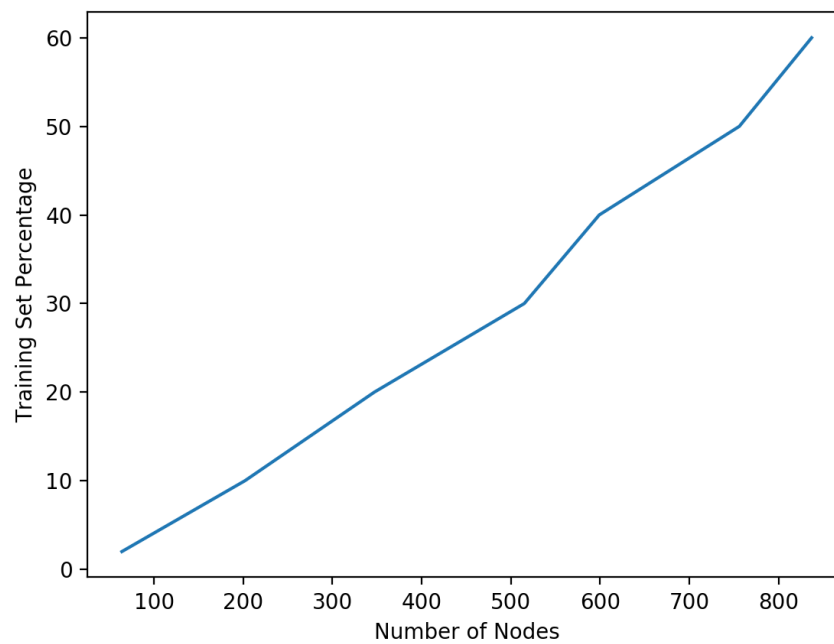
Analysis

1)

Training Set Percentage vs. Test and Train Accuracies

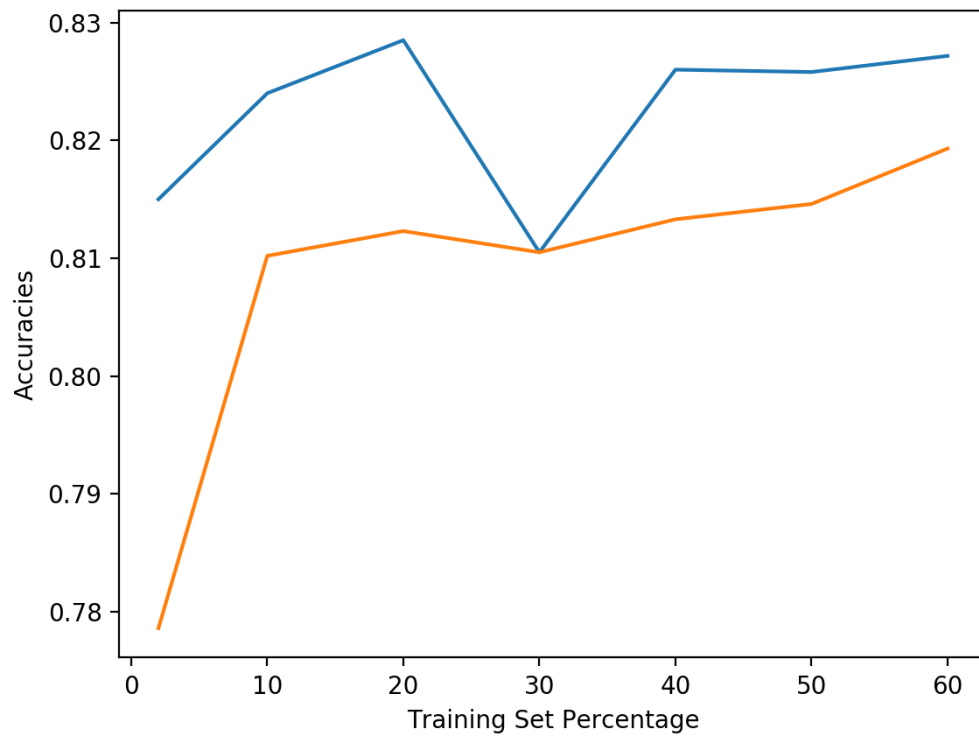


Number of Nodes vs. Training Set Percentage

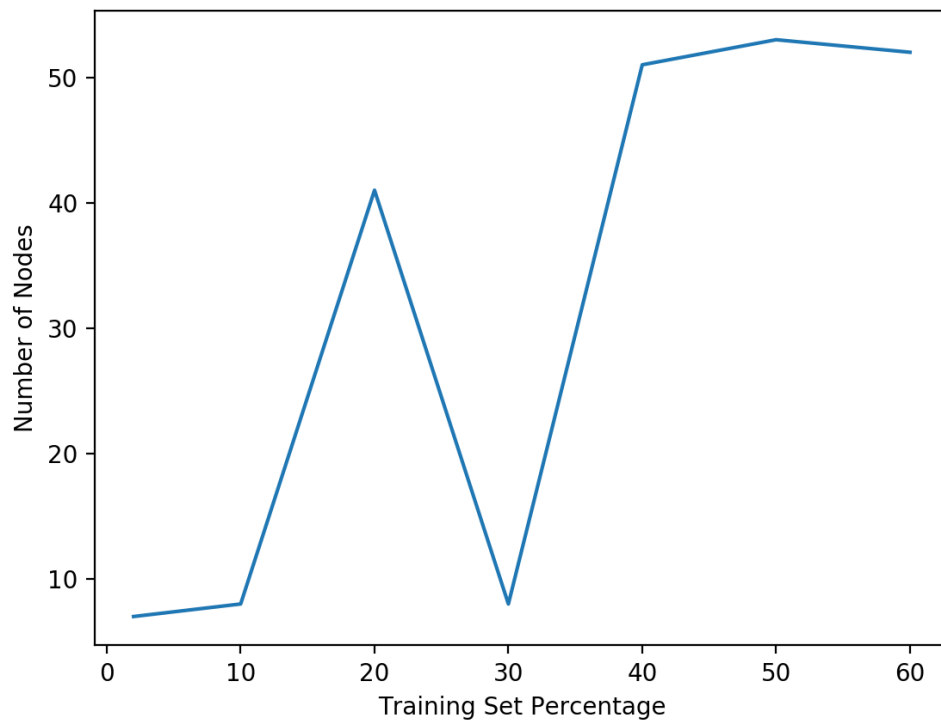


2)

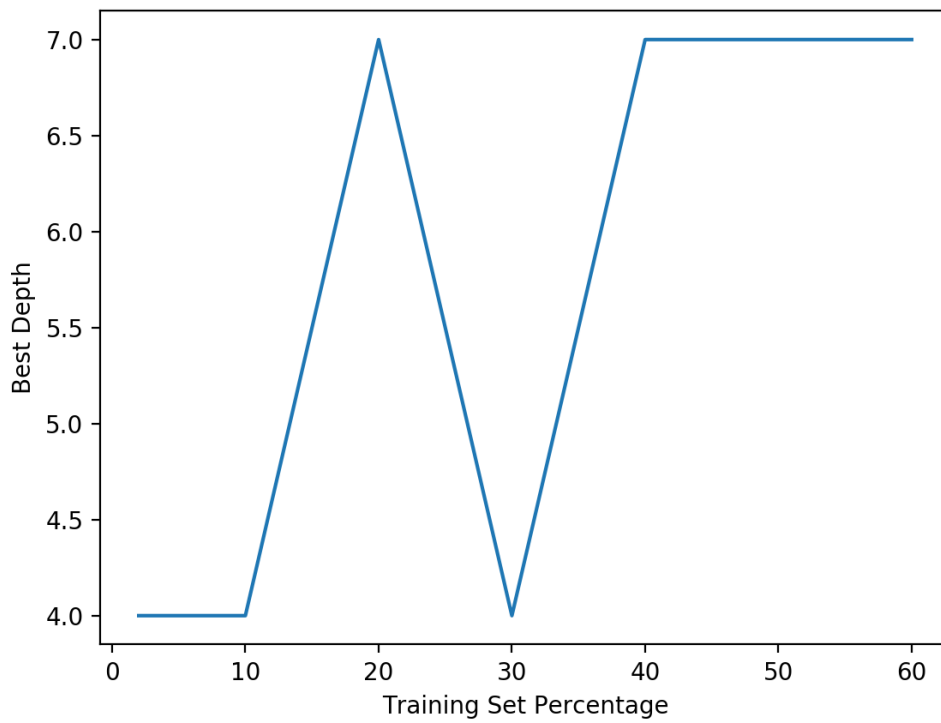
Training And Testing Accuracy with Training Set percentage



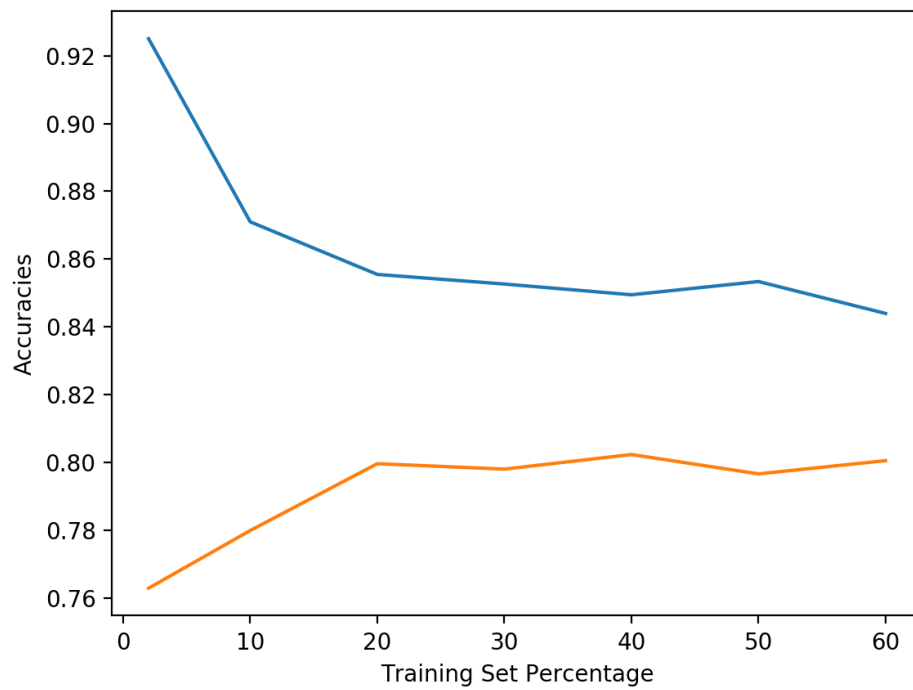
Number of Nodes and Training Set Percentage



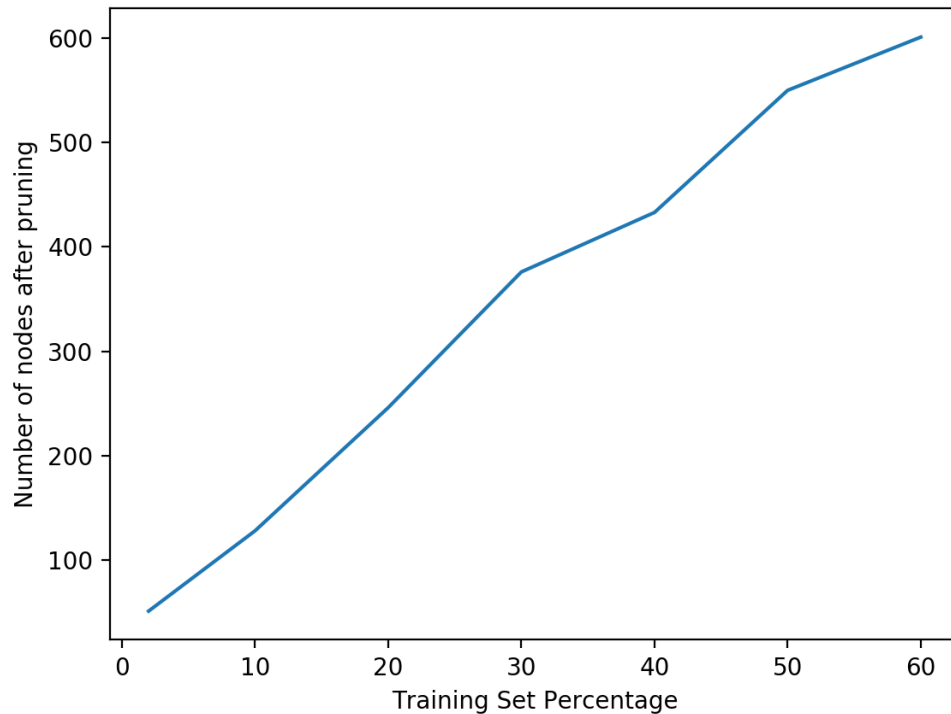
Training Set Percentage and Best Depth for each percent



3) Training and Testing Accuracies and Training Set Percentages



Number of Nodes and Training Set Percentages



4) The test data predictions have to be made using as least biased model as possible. If the tree is pruned on test data, the predictions will be due to an overfit tree. Due to this, the generalization property of the model is highly affected, since the whole point of pruning decision tree on validation set is to increase generalization property by reducing the size of the tree.

5) To convert the decision tree from classification model to ranking model, storing the probabilities in leaf nodes instead of class would be a way. While traversing the tree to find the ranking, the label with the highest probability will be ranked higher than the one with lower probability.