349 HW_1
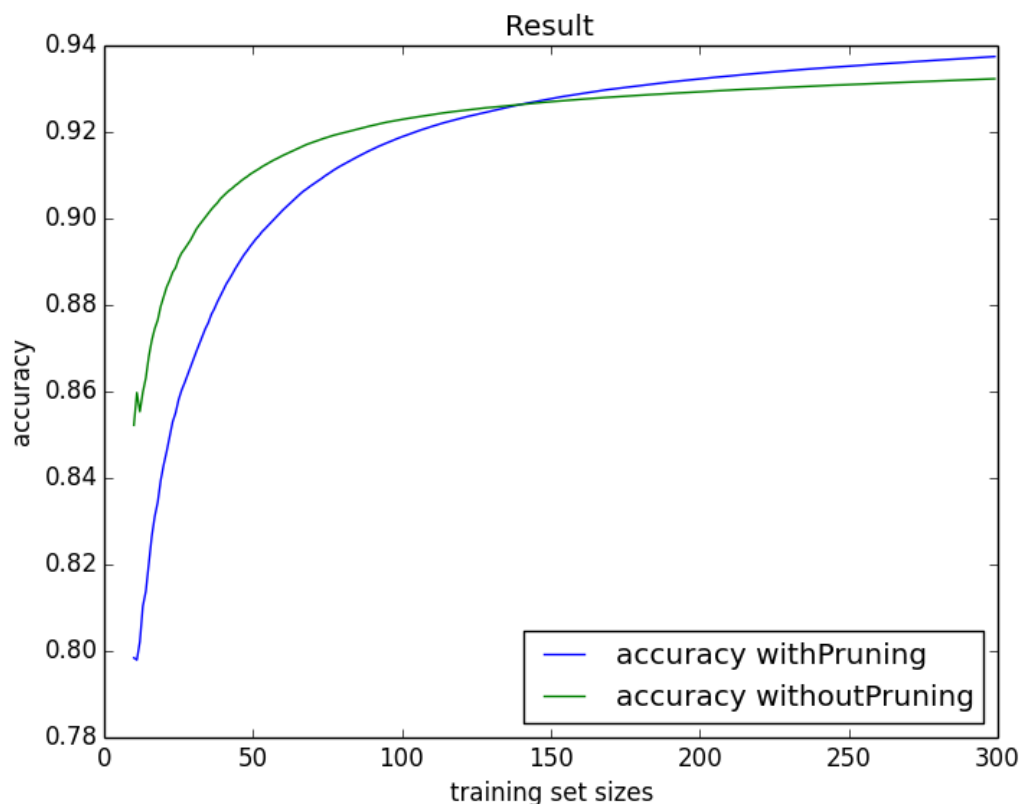
Report

Yunfei Gao

1. My partner is Ting Liu.

2. Yes, we alter the Node data structure. We add MODE attribute to record the most frequent class label in examples at this node. Because when we did out pruning process, we will return the MODE value after cut that node.

3. We treat missing attributes as a new value. We choose this strategy because in this way it can keep features in this dataset to the maximum.

4. We continuously calculate the accuracy if we cut current node. Every time we will cut the node that we can get highest accuracy if we remove it. Do the cut step until we can not improve our accuracy. We choose this strategy because we can get the highest accuracy in every recursive loop.

5.

a. The trend is rising. More training examples can cover more situation, so the tree is

more complete.

b. At first, the result of pruning is worse than without pruning but when the number of training examples increase, the result of pruning is better than without pruning. When number of training examples is small, the tree is not complete, so the result of without pruning is better. When number of training examples is large, the tree has some redundant part, so do some pruning is better.