John Fiedler
CSCE 4201

Chapter 6

1. What is the approximate depth of a Decision Tree trained (without restrictions) on a training set with one million instances?
    a. Since Decision Trees can be thought of as balanced binary trees, the general depth is $log_2(m)$ where m is the number of leaves in the tree. So, if there are one million instances in the training set then the depth of the tree would approximately be $log_2(10^6)$ = 20.
2. Is a node's Gini impurity generally lower of greater than its parent's? Is it generally lower/greater, or always lower/greater?
    a. A node's Gini impurity is generally lower than its parent's.
3. If a Decision Tree is overfitting the training set, is it a good idea to try decreasing max_depth?
    a. Yes, this would be a good idea.
4. If a Decision Tree is underfitting the training set, is it a good idea to try scaling the input features?
    a. It would not help in the case of Decision Trees, so no it would not be a good idea.
5. If it takes one hour to train a Decision Tree on a training set containing 1 million instances, roughly how much time will it take to train another Decision Tree on a training set containing 10 million instances?
    a. I am unsure how to solve this problem.
6. If your training set contains 100,000 instances, will setting presort=True speed up training?
    a. No, it will not speed up training. You would only want to presort the training set if it is relatively small.