Christopher Tsai

CS 349 HW #1 Free Response Questions

1. Decision trees can be used to represent any Boolean function because there is a unique path from every leaf node to the root node. As the amount of Boolean inputs increases, the tree needs at most exponentially as many nodes to represent the function.
2. This majority-rule algorithm can be *represented* by a decision tree because each node only has two options and decision trees can represent any Boolean function. However, it might not be able to *learn* the exact sum-of-all-voters algorithm because the sum-of-all-voters algorithm gives equal decision power to all voters and decision trees naturally give more decision power to some voters (nodes) than others.
3.
    a. candy-data.csv
        i. Accuracy = 0.8
        ii. Number of nodes = 69
        iii. Maximum depth = 9
    b. majority-rule.csv
        i. Accuracy = 1.0
        ii. Number of nodes = 69
        iii. Maximum depth = 6
    c. ivy-league.csv
        i. Accuracy = 1.0
        ii. Number of nodes = 19
        iii. Maximum depth = 4
    d. xor.csv
        i. Accuracy = 1.0
        ii. Number of nodes = 7
        iii. Maximum depth = 2
4. The inductive bias of ID3 is that shorter trees are preferred over larger trees and nodes that have higher information gain are placed closer to the root node of the decision tree.

5. Overfitting means creating a model that works well with only a specific sample of training data; one that does not generalize well to other training samples. In decision trees, one can tell that overfitting has happened if the maximum depth of the tree is very high or if all or most leaves have a single example each.
6. Parallel to the meaning of pruning a real-life tree, pruning a decision tree means getting rid of branches that are unnecessary or redundant to the classification/regression of features. One can either pre-prune or post-prune. Pre-pruning means preventing a tree from growing redundantly by limiting parameters such as num. of leaves, num. of features used, max. depth, etc. as the tree grows. Growth can also be halted if all examples have same class or if all feature values are the same. Post-pruning means pruning after the tree has grown by starting at a bottom non-leaf node and chopping subtrees off, testing the performance of the tree each time and keeping the change if the performance remains unchanged. Post-pruning stops when further pruning hurts performance.
7. When splitting continuous-valued attributes, one first sorts the points by the attribute in question, lowest to highest. Then one calculates the average value of the attribute for all adjacent points (in other words the midpoint between adjacent points). Finally, one calculates the information gain of each midpoint value and picks the value with the highest information gain. This value is the split point. This method works as one is using the metric that is used to build the tree (in this case information gain) to split the values. Gini impurity can also be used instead of information gain.
8. If I were building an ensemble/forest of n trees, I would combine their decisions by taking a majority vote (by using question 2’s algorithm or some other majority vote algorithm such as the Boyer–Moore algorithm). This is how most random forest algorithms make a decision, and it works due to the “wisdom of crowds” (in other words, accuracy increases with number of decision-makers regardless if each individual decision-maker is well-equipped or not).