1. Yes, because the target function is deterministic. The hypothesis space for decision trees is a complete space of all finite discrete-valued functions.
2. A majority-rule algorithm can be represented by a decision tree, but it is not very efficient as the tree would have to learn every possible combination.
3.
    a. Candy
        i. Accuracy: .8
        ii. Number of Nodes: 71
        iii. Maximum Depth: 10
    b. Majority Rule
        i. Accuracy: 1
        ii. Number of Nodes: 19
        iii. Maximum Depth: 5
    c. Ivy League
        i. Accuracy: 1
        ii. Number of Nodes: 69
        iii. Maximum Depth: 9
    d. XOR
        i. Accuracy: 1
        ii. Number of Nodes: 7
        iii. Maximum Depth: 3
4. The ID3 algorithm prefers shorter trees and trees that place attributes with the highest information gain closest to the root.
5. All of the attributes in Majority Rule and XOR have the same information gain, so the tree can pick any of the attributes. Additionally, the tree will have to learn every possible path for both.
6. A hypothesis is overfit if another hypothesis that fits the training data less well performs better. You can tell overfitting is happening when the difference between the accuracy on the training data and the accuracy on testing data begins to increase.
7. Reduced Error Pruning is done by removing branches, then checking the tree's performance over the validation set. If the pruned tree performs no worse, then the cut is kept. Pruning will increase accuracy and decrease the number of nodes in the tree. A drawback is that there will be less training data to learn from because some of it is used for the validation set.
8. First, sort the examples by the attribute. Then, take the average of adjacent examples that have different target classifications. Then, compare the information gain of these averages to pick the split point. This allows us to maximize the information gain.
9. It is possible, but not efficient, because decision trees can only split data parallel to the axis.

10. I would use bagging because it helps reduce the variance of decision trees.

https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/