Question 6:

|  |  |  |
| --- | --- | --- |
| Data Set | Average Classification Rate | Tree Size |
| Dummy1 | 1.0 | 3 |
| Dummy2 | 0.65 | 11 |
| Connect4 | 0.758 | 41521 |
| Car | 0.94675 | 408 |

The Decision tree algorithm had 100% accuracy for Dummy set 1, which makes since sine if we look at the decision tree we see that the 5th attribute determines the class (if it’s a 1 the class is 0, and if it’s a 0 the class is a 1), so this has the highest info gain hence the tree only has size 3. For Dummy set2, the decision tree algorithm does much worse with only 65% accuracy and a tree of size 11. This is likely due to the small training set size (20 examples), so the attributes in the training set might have vastly different info gains than they do in the testing set.

For the Connect4 data set, the decision tree had 75.8% accuracy and a size of 41521. Since there are 42 attributes with 3 values each that 3^42 possibilities. However, there are only 67,557 examples in the training set, thus the tree is very large. The tree does quite poorly since the tree is likely overfitting which could be solved by pruning, since the bottom attributes are likely to be capturing noise in the data set.

For the Car data set, the decision tree had 94.6% accuracy and a size of 408. This data set only has 6 attributes with 3 having 4 values and 3 having 3 values, so that’s 1728 possibilities. The training set had 1728 examples, so potentially there is enough information to classify everything. Since the training set is not too big there is not likely to be overfitting, since there is less noise to capture.

Question 7:

The decision tree for classifying the car data set could be used by used car dealership websites. More attributes can be added such as mileage, price, etc. After a large amount of data is collected, you can run the decision tree algorithm to classify which cars are appealing to certain users. The decision tree can be used to help make recommendations to users.

One could use the minimax algorithm. Since Connect4 consists of 2 players, one can be the maximizer and one the minimizer. The search space is huge so you would probably have to limit how far the maximizer can look ahead. We can use the minimax to decide the attributes to split on for the decision tree.