Q6

Dummy Data Set 1

For the first dummy data set, the final tree has a size of only 3 and a correct classification rate of 1.0.This means that there is some attribute which, for the twenty training examples, allows one to immediately classify the tree correctly. Examining the output from the test, we see that if attribute “5” has a value of zero then we classify as one while if it has a value of one then we classify as zero.

Dummy Data Set 2

The decision tree created from the second dummy data set is unable to classify elements in the testing set as accuratley as in the first, despite having the same number of binary attributes. The key difference is that in this example, there is no perfect attribute to split upon like in the last meaning that the example is not quite so trivial. Comparing the current example count of twenty to the total attribute count of ten, I don’t believe that enough examples are present to form an accurate tree. With the limited example number, our decision tree is running the risk of not learning to split on the best information explaining the relatively low rate of 0.65.

Car Data Set

For the car data set, the decision tree size was 408 and the average classification rate over all runs was 0.923. For any car, the total number of combinations of all of its attribute values is 1728, the same number of examples that it has. The discrepency between the tree size and the attribute value combinations as well as the overall accuracy lead me to believe that there are key attributes that tend to classify the cars very well. Examining car.out shows that safety is one such attribute.

Connect4 Data Set

The connect4 set forms a considerably large tree given its number of examples compared to the car example. I believe this is simply due to there being no outstanding attributes that classify overly well, leading to the average rate being 0.76135.

Q7

Car Example (when do other people make a decision)

A dataset similar to this example might be related to the medical field where symptoms need to be paired with illnesses. A series of symptoms would be the attributes and the classification would be the illness. This would be especially useful as the logical steps can be easily shown to the patient if requested.

Connect4 Example

The decision tree for the connect4 example might perform well as a heuristic function. Because the tree classifies the game as a win, draw, or loss, it would give good representation of how effective a move that put the game into that state would be over others.