Brian Lee P4 Analysis

Question 6:

Dummy Set 1:

Tree size: 3 & Classification Rate: 1.0
There is a perfect classification rate because the 5th attribute perfectly determines the label value of each example. Due to this, the tree size is 3 because only 1 attribute at the root is required to classify examples.

Dummy Set 2:

Tree size: 11 & Classification Rate: 0.65
The classification rate is low because there is a lack of training data. With lack of training data, the decision tree will be inaccurate, because the tree must make classifications based on attributes that may have been deterministic for the small set of training data, but not for testing data. The tree has size of 11 because 20 examples are branched along 5 attributes.

Connect4:

Tree size: 41530 & Classification Rate: 0.757500
The classification rate is low because there exists numerous permutations of possible future moves given any board configuration, thus this nature makes a hard classification task. The tree size itself is huge because there are many permutations of board positions.

Car:

Tree size: 408 & Classification Rate: 0.947750
The classification rate is high because the attributes are informative, discriminating and independent and there is also sufficient data to develop an accurate decision tree. Because there are relatively few deterministic attributes, the tree size is also relatively small.

Question 7:

Cars:
A similar dataset is https://www.kaggle.com/ddmngml/trying-to-predict-used-car-value
Incorporating a decision tree on this dataset will allow us to predict the value of used cars on the market based on attributes such as brand, year of registration, fuel type, etc and help car resellers smartly classify whether the used car is worthy of buying or not.
Connect4:
We can incorporate the classifier with the mini-max algorithm to equip playing bots with the ability to backtrack the game in order to make optimal moves against the human player.

Brian Lee P4 Analysis

Question 8:

Poker Hand Dataset:
Tree size: 399671 & Classification Rate: 0.95425
The classification rate is very high because the attributes are deterministic in predicting the poker hand. However, there is not a perfect classification rate because the decision tree is not trained on all the possible permutations of the poker hands. The tree size is large because there exists numerous possible permutations of future cards even when shown a single card. This makes predicting the final poker hand when not given the entire 5 cards configuration extremely difficult.

We may be able to incorporate value iteration to our decision tree. For every episode of our poker game, when we reach a leaf node with the label values (Royal flush = 10, straight flush = 9, full house = 8…) the values propagate upstream and we can update the expected values of any hand configurations. Later when the bot plays against a human player, the bot can make optimal decisions whether to fold or bet based on the odds that the bot will reach one of the high poker hands in the end.