

Question 6.)

Data Set	Tree Size	Classification Rate
Dummy Set 1	3	1
Dummy Set 2	11	.65
Connect4 Set	41605	.76
Car Set	408	.929

Dummy Set 1 performed well because it is classifiable over a single variable. Variable 5 gives 100 percent information gain.

Dummy Set 2 is less successful because the data set is not neatly classifiable over a single variable this can be seen in the training data. The data is either noisy or more training data is needed to determine a meaningful decision tree from the training data.

The Connect4 Set produces a much larger tree because the data set is far more complex. We are checking over 42 variables. This data set likely overfits because of the large number of examples: 67557, approximately 10% of the data set. If we were to prune this tree we would get a significantly smaller tree with a comparable classification rate.

The Car Set only has 6 variables and trains over 1728 examples. This data set has a really high classification rate indicating that the variables it classifies over are very good indicators. Additionally, the sample size is sufficiently large to classify the dataset.

Question 7.)

The connect 4 decision tree could serve as a heuristic for a connect 4 agent when the agent arrives at a tie. It can serve as a secondary check on the agent as well. If the agent at any point is less certain about its optimal move than the decision tree, then the move will be made based on the decision tree's evaluation.

The cars decision tree can be combined with a car selling application. The buyer can sort based on the variables important to them and the application can narrow down the choices. This approachable is extensible to retailing any multi-faceted product like a house, for example. Applications like Zillow.com, in fact do just that.