Soo Ah Moon
CS3600 Project 4 Analysis

Q6. Performance

The resulting tree size for the Cars dataset was 408, and the average classification rate was 0.9445. The information gained was higher in this set than the Connect4, so the tree size is much smaller in relative to the size of the dataset. The resulting tree size for the Connect4 dataset was 41,521, and the average classification rate was 0.7592. The information gained was overall lower for each iteration compared to the Cars dataset, so the tree size is much larger in relative to the size of the dataset and more random. It was much more difficult to find correlation in attributes and classification for the Connect4 dataset. The resulting tree size for DummySet1 was 3, and the classification rate was 1.0. The resulting tree size for DummySet2 was 11, and the classification rate was 0.65. Due to the small input size of both, they were both more "efficient" when running it as compared to the larger inputs. However, the first dataset had a high correlation with the labels unlike the second dataset. Since the data set (both the training and testing data set) and tree size was small, for any dataset like these, the standard deviation of the accuracy can be very high resulting in a classification rate of 1 keeping in mind that there is a possibility of the first dummy data set having very similar training and testing data sets.

Q7. Applications

The decision tree works like many websites that ask questions (to consumers) about various features (e.g. car features) which continues to narrow down the dataset of products based off of questions/attributes to the "ideal" product for the consumer. Cars websites do this as well as many other online shopping sites. It's possible that many of the "You may also like" suggestions may come from the use of decision trees that agents learn over time to classify.

The Connect4 game classification found made by the Decision tree can be used alongside with value iteration to figure out which spots are best for play on the game board by using the knowledge of how likely the next spot will lead to a win/loss. We can better evaluate a game state as we update the beliefs of our agents based on the opponents' moves as well.