Kenan James

Question 6

The tree generated for the cars dataset yielded a classification rate of about 92.2%, which was better than that of the Connect4 dataset. These results are likely due to the fact that the Connect4 decision tree was composed of attributes don’t correlate highly with the classifications. The dataset is relatively random, and selecting one attribute to split the values is not much more likely than another to cut the size of the remaining subtree. The cars attributes and classifications are much more closely correlated. A look at the examples in cars-data.txt reveals that all the ones with the value “low” for the attribute “safety” received the classification “unacc”, meaning that a care that is unsafe is guaranteed to be unacceptable. Using the safety attribute to split the decision tree allows for large-sized subtrees to be cut out immediately.

Furthermore, the Connect4 decision tree is much larger (size 41,533) than the cars decision tree of size 408. This allows for the introduction of more noise in the data. The fact that the dummy set of size 3 yielded a classification rate of 100% whereas the dummy set of size 11 yielded a 62% classification rate makes this assertion more concrete.

Question 7

With a bit of creativity, the cars data set and decision tree could be used from a business perspective to decide which celebrity should be hired for an advertising campaign for new car models. Provided the company has information on which consumers like which models and what movies they like to watch, the tree could be used to identify which attributes correspond to users that like certain movie genres. That correlation could then be extended to choose a celebrity spokesperson based on which actors star in those movies. An example might have the values vhigh, med, 2, 2, small, low, unacc. Provided that the first value refers to buying price, this combination of values would correspond to a class of expensive cars that are small, require moderate maintenance, and aren’t particularly safe. If this class of cars appeals to a class of consumers who like action movies, particularly those that involve racing, high speed chasing, etc., it might be concluded that the best celebrity spokesperson might be someone like Vin Diesel (who stars in the Fast and Furious movies).

The Connect 4 analysis could be used to determine in which column it is best to drop a piece given a certain game state. Given enough data for an opponent, it could predict where to place the piece based on his/her past actions and the results of those actions. An agent using this tree could possibly learn to classify players as “offensive” or “defensive” and play based on that classification.