

Question 6:

The classification rate of dummy dataset 1 is 100% and the tree size is 3. From the generated decision tree, it's clearly that the classification only depends on the 6th attribute (index 5) of an example. If the 6th attribute (index 5) is true, the result is false. If the 6th attribute (index 5) is false, the result is true. Thus the accuracy is high and the tree size is small.

The classification rate of dummy dataset 2 is 65% and the tree size is 11. Compared to the tree size of dummy dataset 1, that of dummy dataset 2 is much larger. The tree size increases because unlike dataset 1, the classification result does not only depend on one attribute. It means the decision process is more complicated. Because the decision process is more complicated and the data size does not increase (still only 20 examples), the jump of the error rate can be justified.

The classification rates of Connect4 and Car datasets are 76.285% and 94.3% respectively. The tree size of Connect4 dataset is 41521 and the number of examples is 67557. The tree size of Car dataset is 408 and the number of examples is 1728.

Compared to Connect4, the decision tree learning performed extremely well on Car with much higher accuracy and much smaller tree size. I think the success of decision tree learning on Car is because first of all the Car dataset has fewer labels than Connect4, which makes the learning easier. Furthermore, attributes of Car dataset correlate highly with the classification result and classification can be made on top level of the tree in many cases. For example, any low safety car is classified as unacceptable and cars that can only take two people are classified as unacceptable as well.

Connect4 dataset, however, is more chaotic and there does not exist a strong correlation between the attributes and classification results with the added difficulty from more attributes.

Question 7:

For the Car dataset, websites like www.autotrader.com or www.car.com can use the decision tree learning to make suggestion for users. They can collect user inputs like price range, safety rating, trunk size and etc to generate an instance. The website can then apply the pre-learned decision tree from old suggestions on this instance in order to make suggestions. After the suggestion is evaluated by the user, the website knows whether the suggestion is good or not. Then this instance can be incorporated into the dataset to learn the decision tree again.

For the Connect4 dataset, we can incorporate the mini-max decision rule to minimize the possible loss for one action. This way the bot is expected to perform better with more knowledge of an action.