CS 3600 Project 4 Analysis

Jiahao Luo #903103970

Table 1. Results of All Datasets

Dataset	Classification Rate	Tree Size	Number of examples
Dummy dataset 1	100%	3	20
Dummy dataset 2	65%	11	20
Cars	94.10%	408	1728
Connect4	76.15%	41521	67557
Nursery	98.61%	1159	12960

Question 6: Performance

(1) Dummy dataset 1: The tree size is small and the classification rate is high, because the classification only depends on attribute 5. The rest of attributes are irrelevant.

(2) Dummy dataset 2: The tree size is big and the classification rate is low, because the dataset is more complicated and it needs to construct a bigger tree with multiple attributes to classify. Besides, the number of training examples are not enough to fully describe some hidden properties that can affect the classification.

(3) Cars dataset: The tree size is 408 and classification rate is high, 94.10%. The classification rate is reasonable, because all six attributes are essential factors that can affect the values of cars which can greatly help classify different value classes.

(4) Connect4 dataset: The tree size is very large and classification rate is not too hight, 75.15%. Because the final classification result (win, loss or draw) is determined by the combination of all 42 attributes. There is no certain attribute combination pattern that can classify win, loss or draw. The attributes are rarely relevant to the result by themselves.

Question 7: Applications

(1) Cars dataset: The CarMax company has lots of data about the car properties and transaction. When buying cars from customers, then decision tree algorithm can give a better price estimation on cars. This can help avoid paying more money on cars which are not worth that much.

(2) Connect4 dataset: The minimax algorithm can help to make a better connect4 playing bot. We can make a better decision based on utility function according to other players' actions and minimize loss.

Question 8: Novel Dataset

(1) Performance: The Nursery dataset has a very high classification rate and the tree size is very large too. Because all 8 attributes are important features that can influence the recommendation of applications. With significant attributes and a relatively large the number of examples,, the classification rate is reasonable.

(2) Applications: The Nursery dataset can be used to rank applications for nursery schools and can give rejected applications an objective explanation.