

#### Problem 6:

Classification rate of dummy set 1: 100%, Treesize:3, sample:20

The process is effective and simple. The decision is only determined by attribute 5. So the tree is small.

Classification rate of dummy set 2: 65%, Treesize:11, sample:20

The process is ineffective. Too many attributes are concerned, so the Treesize is large. But the sample size of the training set is not large enough to provide the training.

Classification rate of connect4: 75.64%(average), Treesize:41521, sample:67557

The process is not so effective. The connect4 samples have too many attributes that make the classification hard and treesize large. Moreover, the attributes are somewhat random and not so logically related to the result and lower the classification rate.

Classification rate of Car dataset: 94.475%(average), Treesize:408, sample:1728

The process is effective. Car dataset samples have proper attributes. The attributes are also meaningful and logically related to the result that lift the successful classification. For example, nearly no one want to buy low safety car and the car only contains two people.

Classification rate of tic-tac-toe(extra credit) dataset: 86.9%(average), Treesize:376, sample:958

Resource: <http://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/tic-tac-toe.names>

<http://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/tic-tac-toe.data>

The process is effective. Tic-tac-toe dataset samples have proper attributes (only 9). It is also a dataset from game, like connect 4, so the attribute value is also somewhat random and not guaranteed to be logical coherent. But since it only has 9 attributes, much smaller than connect4(which has 42), it performs better.

#### Problem 7

Car dataset:

It can help online car sellers to recommend car to the customer according to their preference. Customer can select whether they want a car with 2-4 doors, or a car has space for 2-more people, or a car with low or high maintenance fee or a cheap or expensive car or a car of some specific color. The agent to recommend car to customers according to their requirement. The agent can also record and learn from costumers' final choice and their preference, so that continue refining the dataset and decision tree.

Connect4:

By extensive learning from larger dataset, the game agent will know heuristics for where the "power spot" is on the chessboard from the result of the decision tree. And we can use minimax or expectimax algorithm to evaluate the game state.

Tic-tac-toe(extra):

The dataset is also a fetched out from a game. The application and improvement is similar to that stated in connect4.