Machine Learning : Decision Trees

Property of dataset

Dataset	***Condition for true classification***
MONK-1	(a{1} =a{2}) OR (a{5} = 1)
MONK-2	a{i}= 1 for exactly two i in {1,2,...,6}
MONK-3	(a{5} = 1 AND a{4} = 1) OR (a{5} = 1 AND a{4} = 1)

Assignment_0 :

We can conclude that the MONK-2 is harder to learn because the height of the decision tree is much higher and the entropy of the data is bigger due to many conditions to check before classifying the dataset as true or false. Furthermore, the computational complexity to decide how to classify is much costly due to lack of a good attribute to split up the dataset in two pure classes true and false.

Assignment_1 :
Calculation of the entropy of each dataset

Dataset	***Entropy***
MONK-1	1.0
MONK-2	0.957117428264771
MONK-3	0.9998061328047111

Assignment_2 :

The entropy is a calculation from information theory and it is a cornerstone method to calculate the degree of impurity or heterogeneity of a set. A uniform distribution distribution is distribution there the probability of each outcomes is the same, like the binomial distribution. This kind of distribution has a very high entropy because independent of what the input is the outcome will be the same, it is foreseeable. In non-uniform distribution, different input may give different outcomes, and you cannot in advance know for sure what the outcome could be. That is the entropy is very low, like the normal distribution.

Assignment 3 :
Calculation of the information gain for each attributes

Dataset	*a1*	*a2*	*a3*	*a4*	*a5*	*a6*
MONK-1	0.07527	0.0058	0.004707	0.02631	0.28703	0.000757
MONK-2	0.003756	0.00246	0.0010561	0.01566	0.017277	0.00625
MONK-3	0.007121	0.293736	0.000831	0.002892	0.255912	0.007077

Assignment 4 :

When the overall information gain of the dataset is maximized, the Sk set must be very small and the size of the |S| must be big and Sk become small as possible.
We can argue that the information gain is a good heuristic for picking an attribute for splitting because when we choose the attribute with the biggest information gain the impurity of set become small and uncertainty low that means we can fast determine how to classify input data with some given attributes.

Assignment 5 :

Dataset	E-train	E-test
MONK-1	1-1.0	1-0.8287 = 0.18
MONK-2	1.0	0.6921
MONK-3	1.0	0.9444

Assumptions : My assumptions was that the E-train should perform 100% correctly.
E-train perform 100% correctly because the tree is build based on the training data, but we see that the MONK-2 data has the biggest fraction of misclassified input, that was exactly our assumptions about the difficulty to learn the MON-2 dataset. The dataset have a tendency to overfitting.

Assignment 6 :

if pruning is used to minimize the error, then it can be used to diminish the variance and control by reducing the depth of the tree. The overfitting happens when the model learn from noise data, and bias is when erroneous assumptions occurs and the miss to find a relevant relationship between the features and target outputs that may lead to underfitting. Pruning can help reducing the risk to overfitting by reducing the variance hence the depth of the generated tree.

Assignment 7 :