Assignment0:

MONK1 the dependency level between the attributes exists but it is not very strong. Only three attributes that influence the decision.
MONK-2 all attributes are independent, a depth of 6 is required. Therefore, this dataset is assumed to be the most difficult to learn.
MONK-3: This dataset contains noise according to the problem description that states the fact that the dataset has 5% additional noise in the training set. The attributes dependence level is highest in this data set. 4 different attributes influence the result. Should be the easiest out of the 3 datasets to learn.

Assignment1:

$$\text{Entropy}(S) = -p_0 \log_2 p_0 - p_1 \log_2 p_1 \tag{2}$$

where p_0 and $p_1 = 1 - p_0$ are the proportions of examples belonging to class 0 and 1.

Dataset	Entropy
MONK1	1
MONK2	0.95712
MONK3	0.99981

Assignment2:

Uniform distribution:
Entropy is the measure of uncertainty, and therefore will be maximized if the domain is uniform. Maximum entropy corresponds to the least amount of knowledge about the domain.
Fair six-sided die, Fair coin

Non-uniform distribution:
The entropy for a non-uniform distribution will be lower than for the uniform one. This is because a non-uniform distribution is easier to predict and thus have a lower amount of uncertainty.
Biased coin

Assignment3:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{k \in \text{values}(A)} \frac{|S_k|}{|S|} \text{Entropy}(S_k) \qquad (3)$$

Dataset	a1	a2	a3	a4	a5	a6
monk1	0.07527	0.00584	0.00471	0.02631	0.28703	0.00076
monk2	0.00376	0.00246	0.00106	0.01566	0.01728	0.00625
monk3	0.00712	0.29374	0.00083	0.00289	0.25591	0.00708

Assignment4:

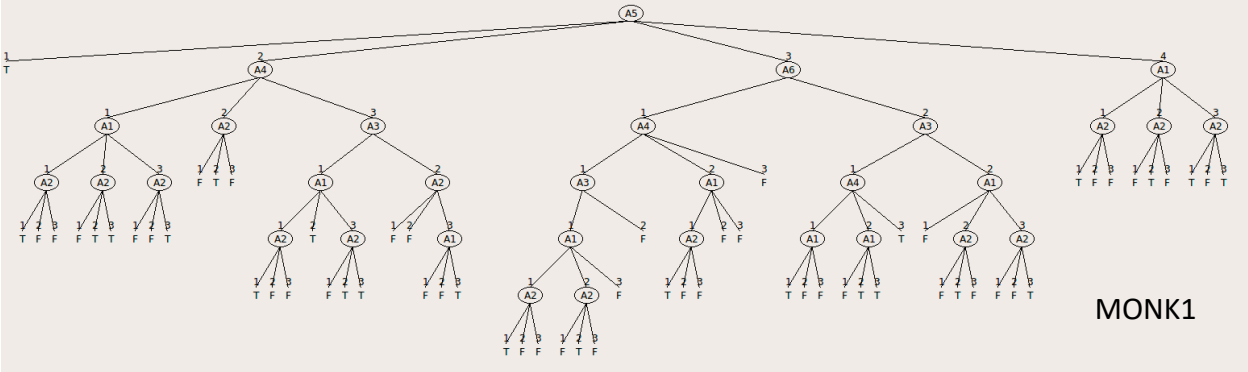
When the information gain is maximized, that means more information is acquired and thus the entropy is decreasing.

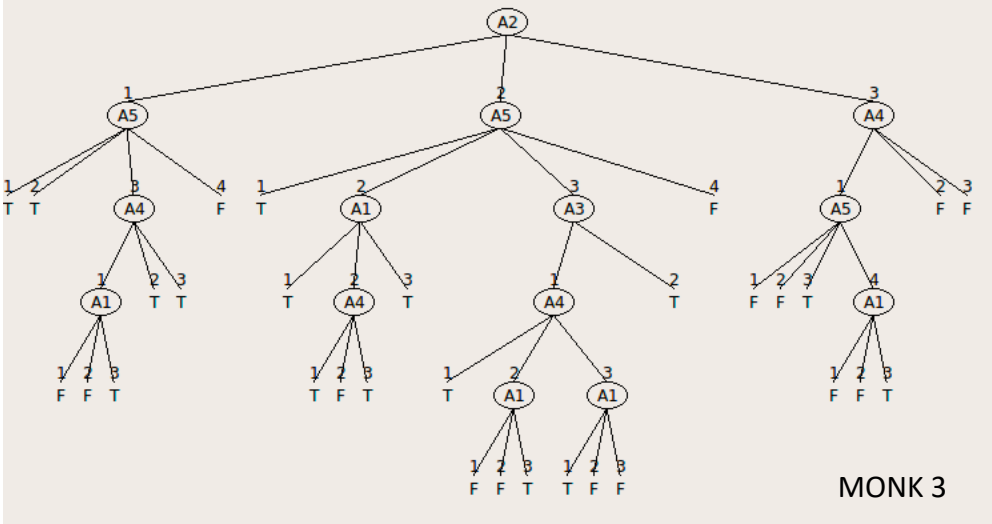
Selecting the attribute with the largest information gain would decrease the uncertainty (entropy) of the subset the quickest. As a result, it is expected that lesser subsets are required, and in return produce the smallest tree (increasing performance).

Assignment5:

Dataset	Error-train	Error-test	
MONK1	0	0.1713	2
MONK2	0	0.3079	3
MONK3	0	0.0556	1

MONK-3 error results agreed with what was expected. The dataset included a good amount of dependency between its attributes. the 5% additional noise did not affect the accuracy as much as expected.
MONK-2 error result agreed with what was expected. All the attributes were independent, and it would require a depth of 6. Even the biggest gain value was very small compared to MONK-1 and MONK-2
MONK-1 did not have attribute dependencies as large as Monk-3, so it was expected not to perform as good as MONK-3.





Assignment6:

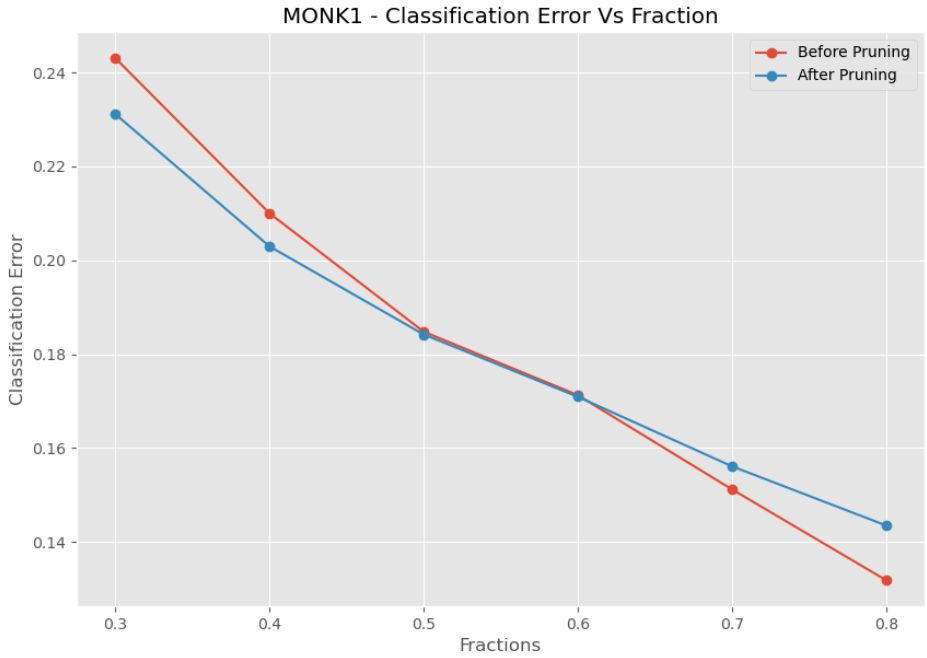
Pruning reduces the model complexity, as the size of the decision tree is reduced. Decision trees often has a high variance and low bias, due to the overfitting nature of the model. The variance will be reduced by pruning, and as a result bias is increased.

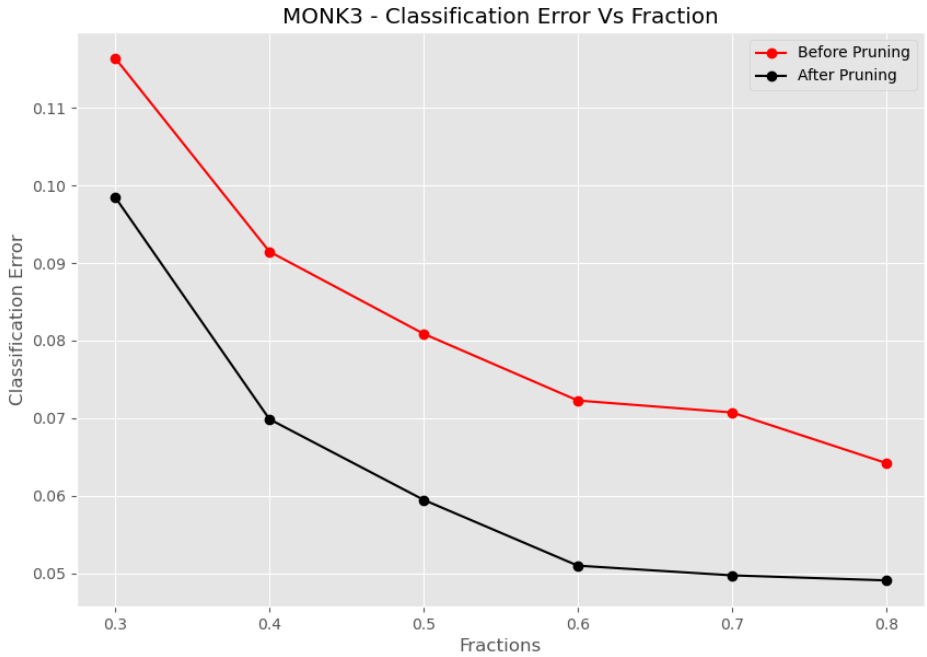
Assignment7:

With a larger fraction value (using more data for training) results in a smaller classification error. The results also showed that MONK-3 has benefited a lot more from pruning compared to MONK-1.

-0.8 fraction for monk-1 showed best result for classification error

-0.7 fraction for monk-3 showed best result for classification error


MONK1 - Classification Error Vs Fraction
Before Pruning
After Pruning
Classification Error
0.24
0.22
0.20
0.18
0.16
0.14
0.3
0.4
0.5
0.6
0.7
0.8
Fractions


MONK3 - Classification Error Vs Fraction
Before Pruning
After Pruning
Classification Error
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.3
0.4
0.5
0.6
0.7
0.8
Fractions