# Assignment0:

**MONK1** contains both attribute a1 and a2 in the same relation, which means they will be difficult to split. Fairly few data points in training set (124). Only three attributes that influence the decision, which means only a depth of 3 should be required to decide all outcomes.
**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. Since only three attributes influence the decision, a depth of 3 is enough to decide all outcomes. Should be the easiest out of the 3 datasets to learn.

# Assignment1:

| Dataset | Entropy |
|---|---|
| MONK1 | 1 |
| MONK2 | 0.95712 |
| MONK3 | 0.99981 |

# Assignment2:

**Uniform distribution:**
Entropy will be maximized if p is uniform. maximum entropy corresponds to the least amount of knowledge about the domain d, which is the uniform distribution.
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:

| 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:

If we pick the attribute with the largest information gain, that means entropy will be reduced the most by this attribute. This also means the attribute is more important than others in the decision making.

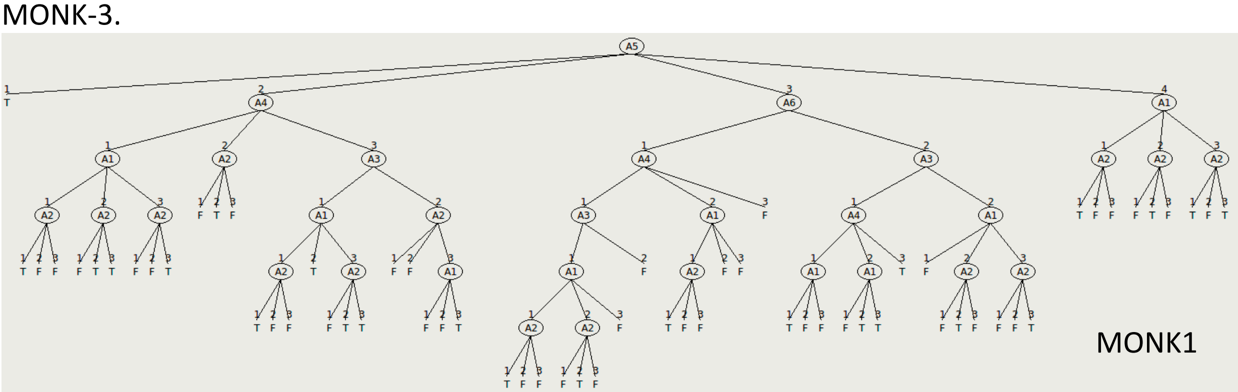
When the information gain is maximized, that means more information is acquired and thus the entropy is decreasing.

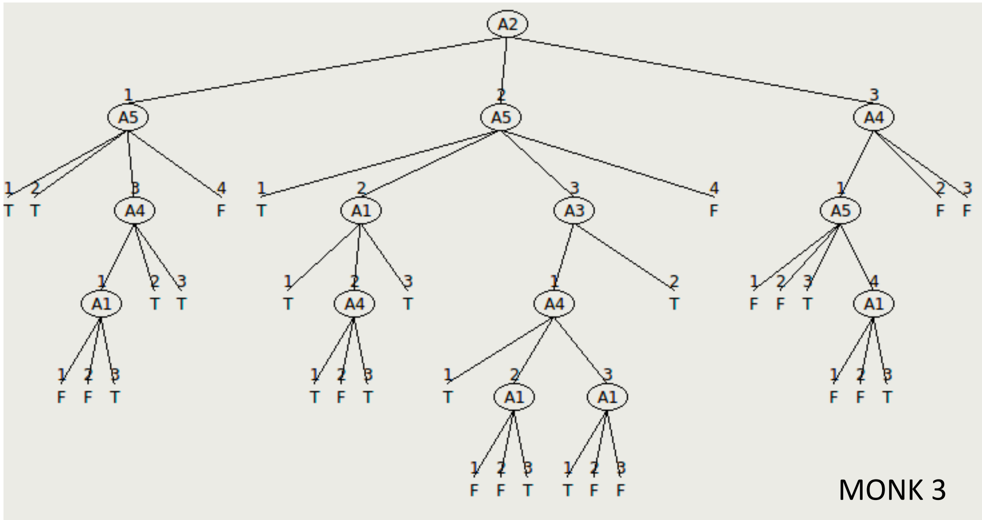
If we use the information gain as a heuristic we want to maximize when making a split, that will ensure selecting an attribute to split on, which reduces the uncertainty (entropy) of the subsets the most. Followingly, this reduces the number of splits and subsets - i.e. results in the smallest tree with the purest subsets, which increases the performance of the decision tree algorithm. An attribute with a high information gain, implies a branch which is more homogenous and predictable.

# 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. It didn't include any attributes that are equal to each other. 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 had 2 attributes equal to each other, so it was expected not to perform as good as MONK-3.



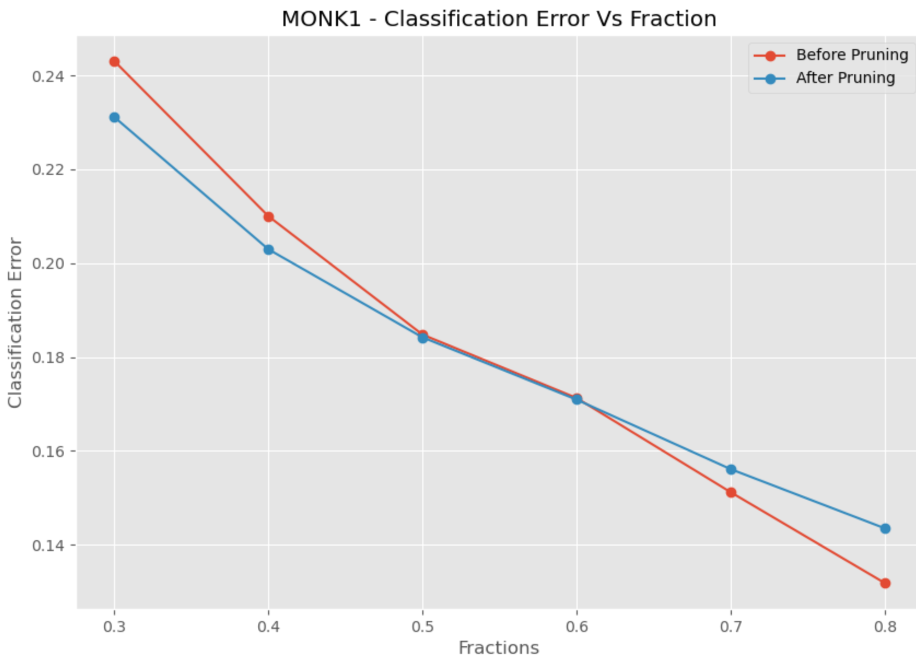


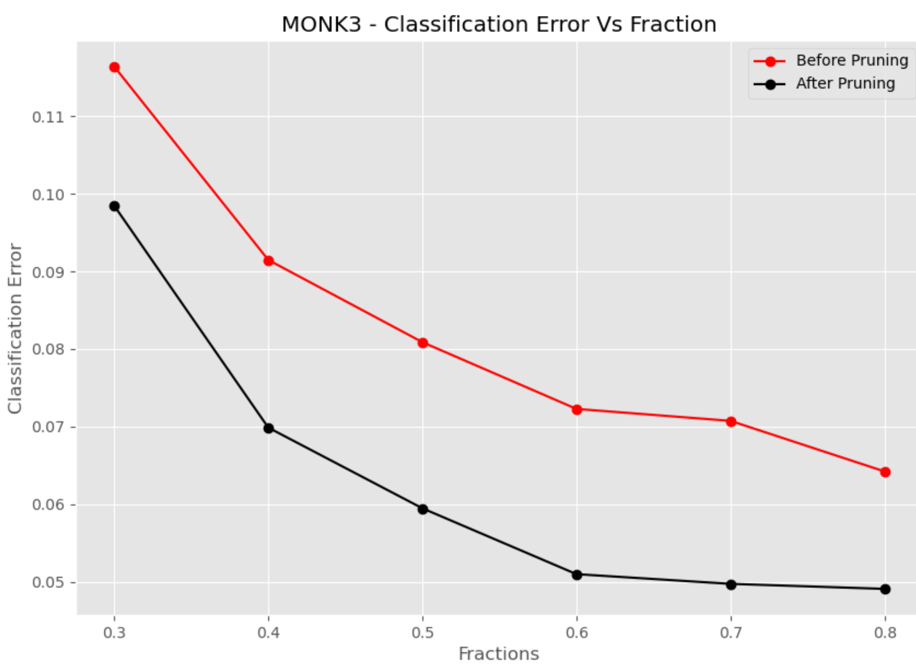

# Assignment6:

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 a result bias is increased. This also makes the model more general and usable.

# 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.


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


MONK3 - Classification Error Vs Fraction
Before Pruning
After Pruning
Classification Error
Fractions
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