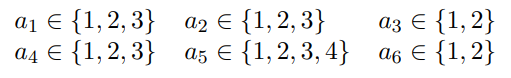
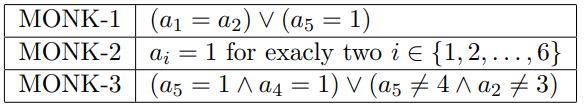
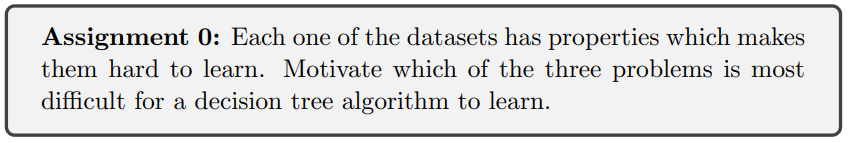
Lab 1 – Decision Trees

# MONK Datasets:



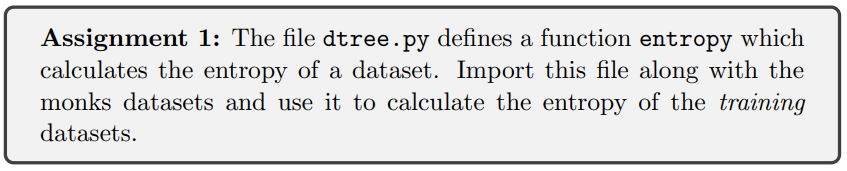
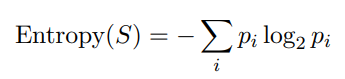


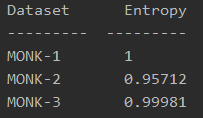


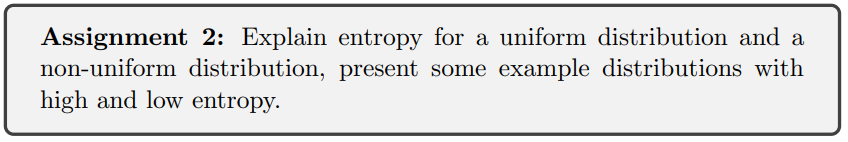
MONK-2 is the dataset which is the hardest to learn. This is due to each outcome is an individual case where each attribute is relevant and needs to be checked. E.g. when a1=a2=1 still requires all other attributes to be checked.

MONK-1 and MONK-3 on the other hand have less relevant attributes and more shared outcomes. This leads to a less complex tree and is therefore easier for a decision tree algorithm to learn.

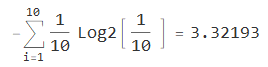
# Entropy







Entropy is a measurement of uncertainty. It increases when it becomes harder to predict the outcome of an event, e.g. throwing a 10-sided dice:



It decreases when the certainty of an outcome increases. When there’s only one possible outcome the entropy is 0.

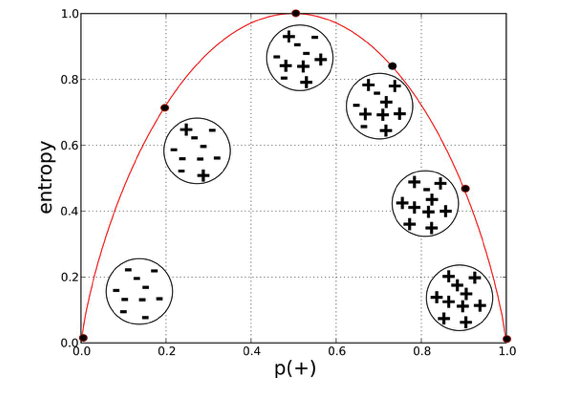
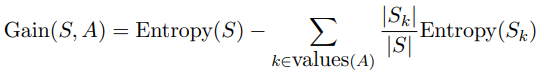
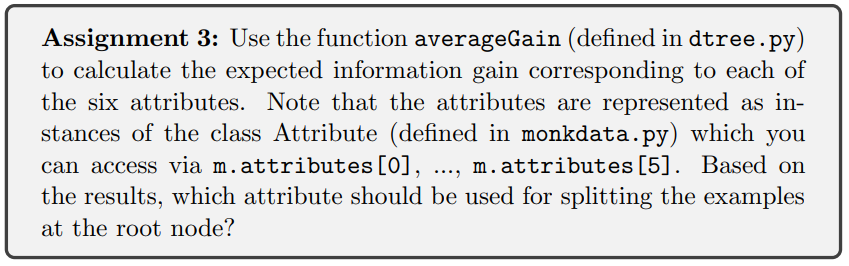
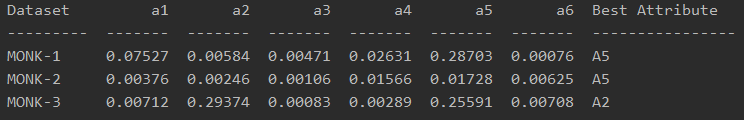


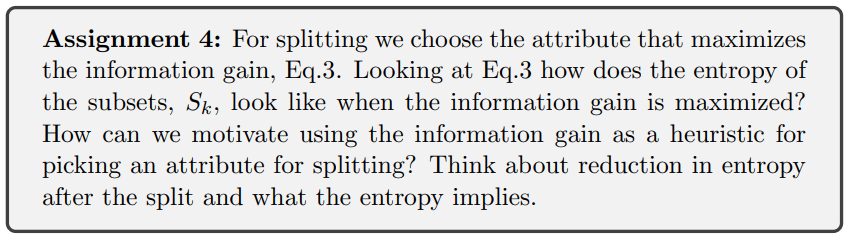
Figure 1. Shows entropy depending on the distribution of "+" and "-".

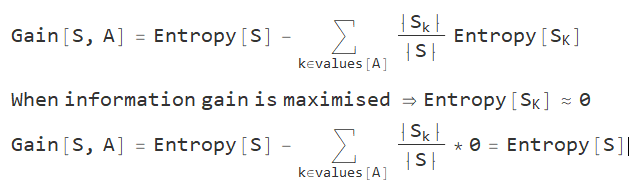
# Information Gain





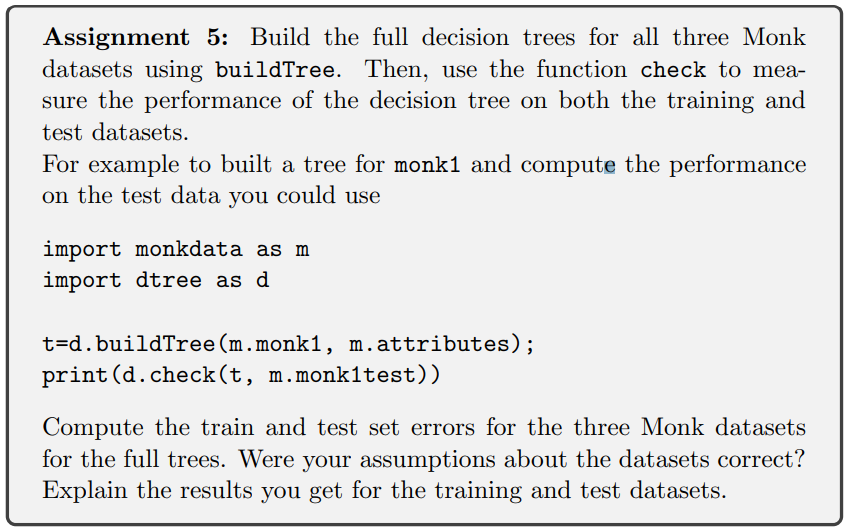


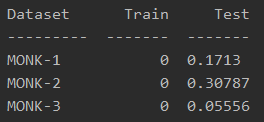




To minimise the complexity of a tree the *most effective questions* should be asked first, which is in this case are the attributes which hold the most information. Therefore, they should be the first nodes in the tree.

# Building Decision Trees

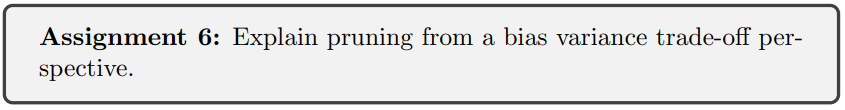




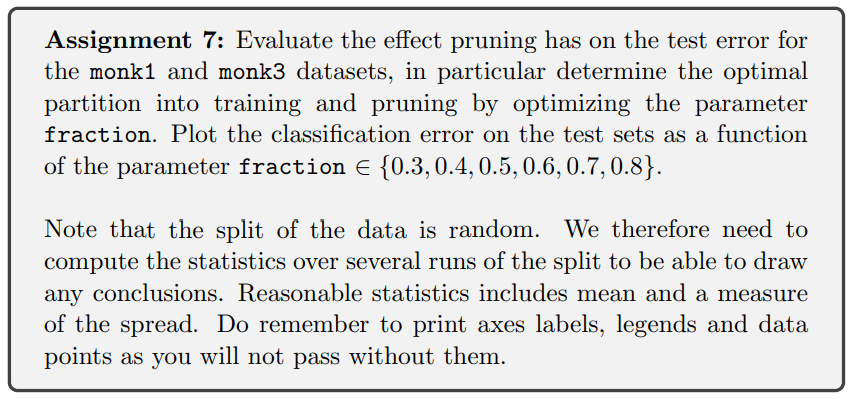
The training dataset was used to generate the trees, so it is not surprising that they didn’t have any errors. MONK-2 was previously mentioned to have been the more complicated dataset which also explains why more errors are present in the test set.

We can see that overfitting occurs because there are no errors in the train set but errors in the test set. Introducing some noise in a dataset can help to reduce overfitting and make the decision tree more generalized which might explain the good result from MONK-3.

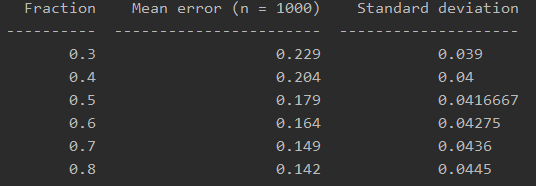
# Pruning

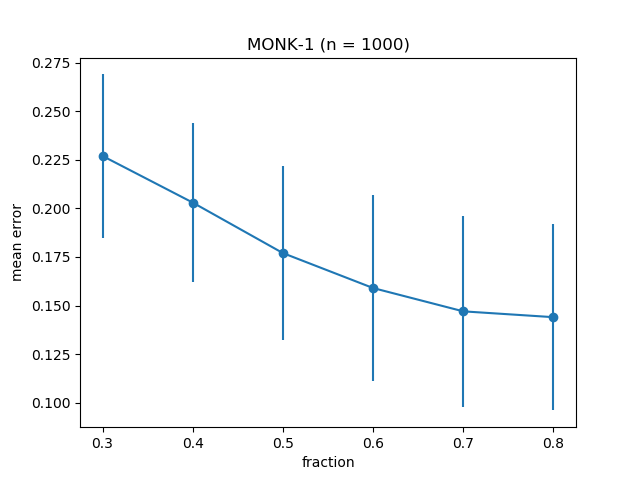


The depth/complexity of a decision tree determines the variance. by pruning we're simplifying the tree and therefore are reducing the variance (reduces overfitting). But by reducing the variance the bias increases, so a trade-off must be made.

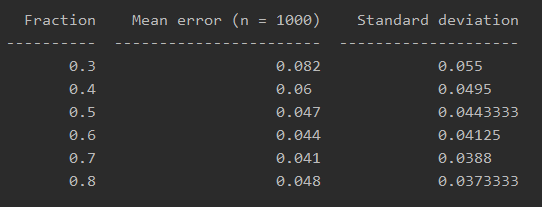


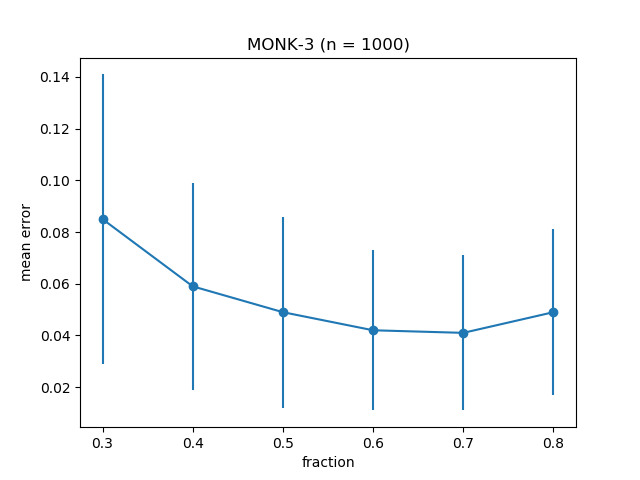
**MONK-1:**





**MONK-3:**



****