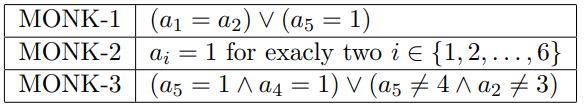
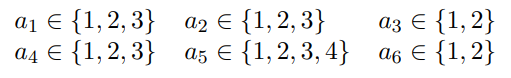
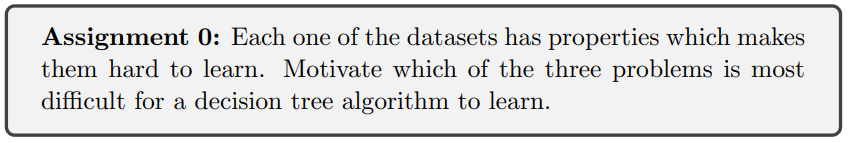
# ​1.​ Preparations

...

# ​2.​ MONK datasets

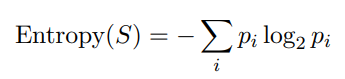


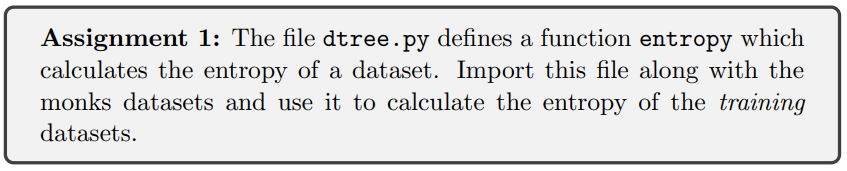


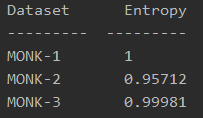
MONK-2 is the dataset which is the hardest to learn. This is due to each outcome is an individual case where every attribute is relevant and needs to be checked. E.g. if a1=a2=1 still requires that a3 to a6 are checked in the tree structure.

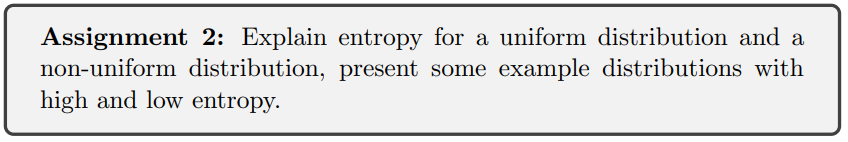
MONK-1 and MONK-3 on the other hand have less relevant attributes and more shared outcomes. This leads to a less complex tree and is therefore easier for a decision tree algorithm to learn.

# ​3.​ Entropy

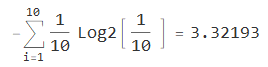








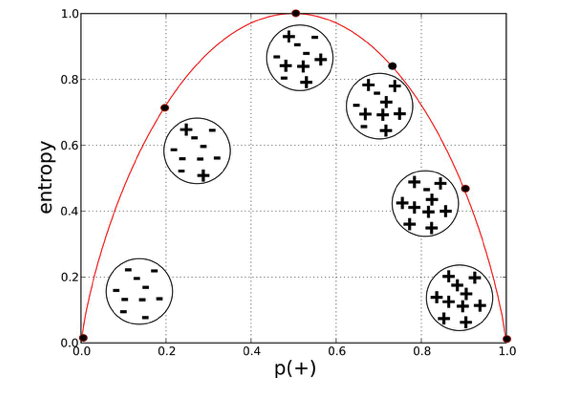
Entropy is a measurement of uncertainty. It increases when it becomes harder to predict the outcome of an event, e.g. throwing a 10 sided dice:



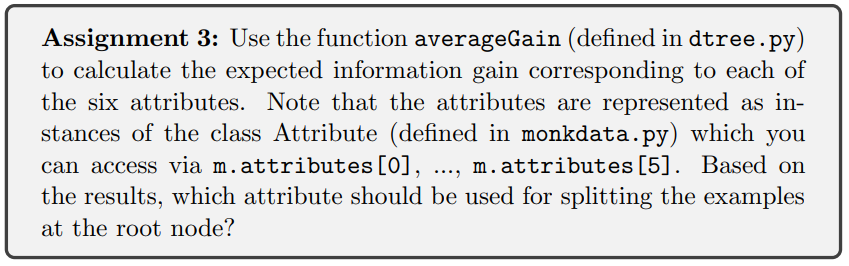
Entropy decreases when a specific outcome of an event increases, e.g. a coin toss of an uneven coin where heads has a likelihood of 0.8 and tails 0.2:

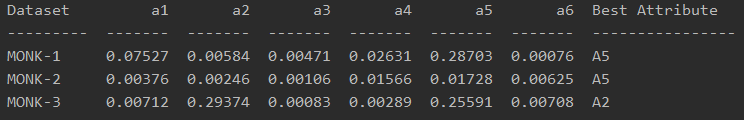


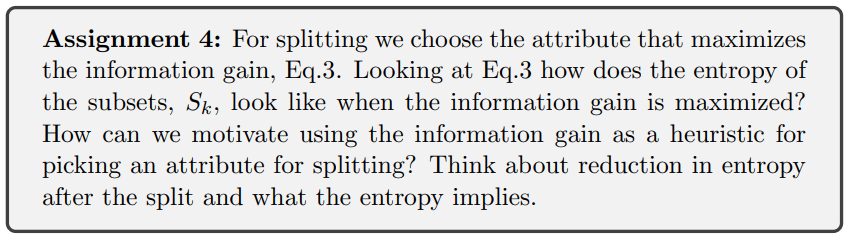
The following figure shows the relationship well:

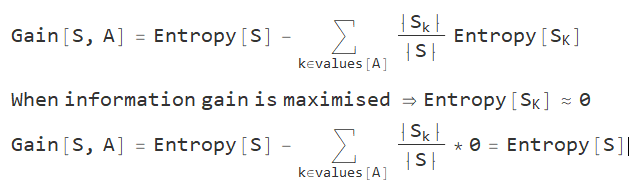


# ​4.​ Information Gain



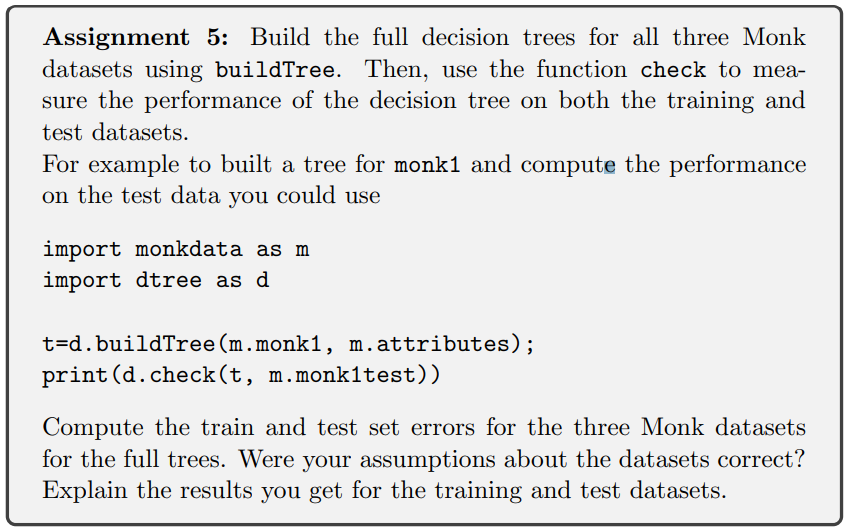


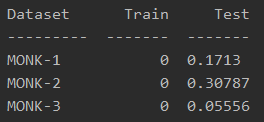




To minimise the complexity of a tree the *most effective questions* should be asked first. Which is in this case are the attributes which hold the most information, and they should be first in the tree.

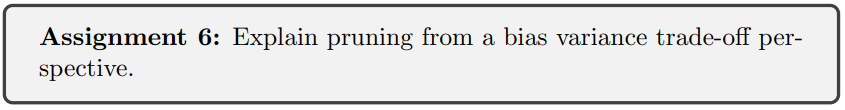
# ​5.​ Building Decision Trees





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. Did some overfitting occur for MONK-1? When a few samples were removed the error rate decreased by roughly 0.05.

# ​6.​ 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 has to be made.

