**Assignment 0**

The difficulty of learning depends on the number of questions to be asked to decide. So the question is how time consuming the training of the decision tree is. If the training data is noisy the training of the true model is more difficult since there a false-true values and vice versa.

* MONK-1: Decision possible after 2 boolean questions.
* MONK-2: Decision possible after 6 boolean questions.
* MONK-3: Decision possible after 4 boolean questions.

**Assignment 1**

Entropy for the MONK Data

* MONK-1: 1.0
* MONK-2: 0.9571
* MONK-3: 0.9998

**Assignment 2**

For uniform distribution the probability of an event is the same for every event. .

A typical example is a fair dice where the probability of every event is .

For a nonuniform distribution the probability at least for one event differs. So, the sum of every possible event is needed.

An example for a nonuniform distribution is an unfair dice where one number occurs with a higher probability, e.g. . As a result, the entropy decreases since the information decays because it’s likely to predict the outcome.

**Assignment 3**

The information gain is calculated by the given script. The following output is generated.

Information Gain Monk 1

* Attribute 1: 0.07527255560831925
* Attribute 2: 0.005838429962909286
* Attribute 3: 0.00470756661729721
* Attribute 4: 0.02631169650768228
* Attribute 5: 0.28703074971578435
* Attribute 6: 0.0007578557158638421

Since Attribute results in the highest information gain it should be used to split at the root node.

Information Gain Monk 2

* Attribute 1: 0.0037561773775118823
* Attribute 2: 0.0024584986660830532
* Attribute 3: 0.0010561477158920196
* Attribute 4: 0.015664247292643818
* Attribute 5: 0.01727717693791797
* Attribute 6: 0.006247622236881467

Since Attribute 5 results also in the highest information it should be used to split at the root node.

Information Gain Monk 3

* Attribute 1: 0.007120868396071844
* Attribute 2: 0.29373617350838865
* Attribute 3: 0.0008311140445336207
* Attribute 4: 0.002891817288654397
* Attribute 5: 0.25591172461972755
* Attribute 6: 0.007077026074097326

Since Attribute 2 results in the highest information gain the tree should be splitted with it as a root node.

**Assignment 4**

For the attribute of the highest information gain the is reduced the most. So, this subset has the lowest weighted sum of entropies. So the most lowest uncertainty.