Assignment 0

MONK-1	$(a_1 = a_2) \vee (a_5 = 1)$
MONK-2	$a_i = 1$ for exacly two $i \in \{1, 2, \ldots, 6\}$
MONK-3	$(a_5 = 1 \wedge a_4 = 1) \vee (a_5 \neq 4 \wedge a_2 \neq 3)$

- **The hardest one is Monk2**

- Monk1: a3, a4 and a6 don't have influence on the decision. Only 4 possible combinatione, therefore the tree only needa a depth of three to decide all the outcomes
- Monk2: all the attributes are independent, all the possible combinations are needed. Depth of six
- Monk3: As monk1 need only depts of three, but it hase the misclassification noise, which makes it harder to read.

Assignment 1

- **Entropies** of the three datasets:

DATASET	ENTROPY
MONK-1	1.0
MONK-2	0.9561
MONK-3	0.9998

Assignment 2

- **Uniform distributions have higher entropy** since all all the datapoint are piscked with equal probability, hence the randomness increases
- **Non-uniform distributions have lower entropy** since the randomness is lower
- Example the entropy of a normal dice (uniform) is equal to 2.58, which is higher than the entropy of a fake dice (non-uniform) that is 2.16

Assignment 3

- The attribute that should be used for splitting at the root node is the one with the **higher information gain,** since an higher information gain means an higher entropy reduction after the splitting.

Dataset	a1	a2	a3	a4	a5	a6
MONK-1	0.07527	0.00584	0.00471	0.02631	0.28703	0.00076
MONK-2	0.00376	0.00246	0.00106	0.01566	0.01728	0.00625
MONK-3	0.00712	0.29374	0.00083	0.00289	0.25591	0.00708

Assigment 4

- Looking at the equation, the entropy of the subset S_k decrease when the information gain is maximized
- When the information gain increases, the entropy decreases
- By picking the attribute which maximize the information gain, we will have the higher entropy reduction.
- We know that a low entropy implies lower randomness in the dataset, i.e. more certainty about the classification

Assignment 5

- Errors for the three full trees

Datasert	E_{train}	E_{test}
MONK-1	0	0.1713
MONK-2	0	0.3079
MONK-3	0	0.0556

Assignment 6

- By pruning a tree we reduce its complexity. This implies lower variance and higher bias than the original tree
- Pruning too much will implies to have a too simple model, which will have a too high bias
- We must find a trade-off

Assignment 7