MACHINE LEARNING
LAB 1 – DECISION TREES

ASSIGNMENT 0

Dataset	True concept
MONK-1	$(a1 = a2) \vee (a5 = 1)$
MONK-2	$a_i = 1$ for exactly two i $\in \{1, 2, \ldots, 6\}$
MONK-3	$(a5 = 1 \wedge a4 = 1) \vee (a5 \neq 4 \wedge a2 \neq 3)$

MONK-2 is the most difficult dataset to learn because it requires a global logic instead of a local one, but decision trees are usually trained based on single attributes and their information gain or GINI index.

ASSIGNMENT 1

Dataset	Entropy
MONK-1	1.0
MONK-2	0.9571
MONK-3	0.9998

ASSIGNMENT 2

In general:

- Uniform distribution → min predictability → highest entropy (With N classes $\log_2 N$)
- Non-uniform distribution → higher predictability → lower entropy

With 2 classes:

- Uniform distribution → entropy = 1
- Non-uniform distribution → entropy < 1

The extreme case opposite to the uniform distribution is when all the samples belong to the same class. In this case, the entropy is 0 because we have the highest possible predictability.

ASSIGNMENT 3

Dataset	a1	a2	a3	a4	a5	a6
MONK-1	0.0753	0.0058	0.0047	0.0263	0.287	0.0008
MONK-2	0.0038	0.0025	0.0011	0.0157	0.0173	0.0062
MONK-3	0.0071	0.2937	0.0008	0.0029	0.2559	0.0071

ASSIGNMENT 4

- Using the split that maximizes the information gain, the entropies of the subsets are such that their weighted sum is minimized.

- This implies that the predictability after splitting is improved and the improvement is the best among the possible splits.

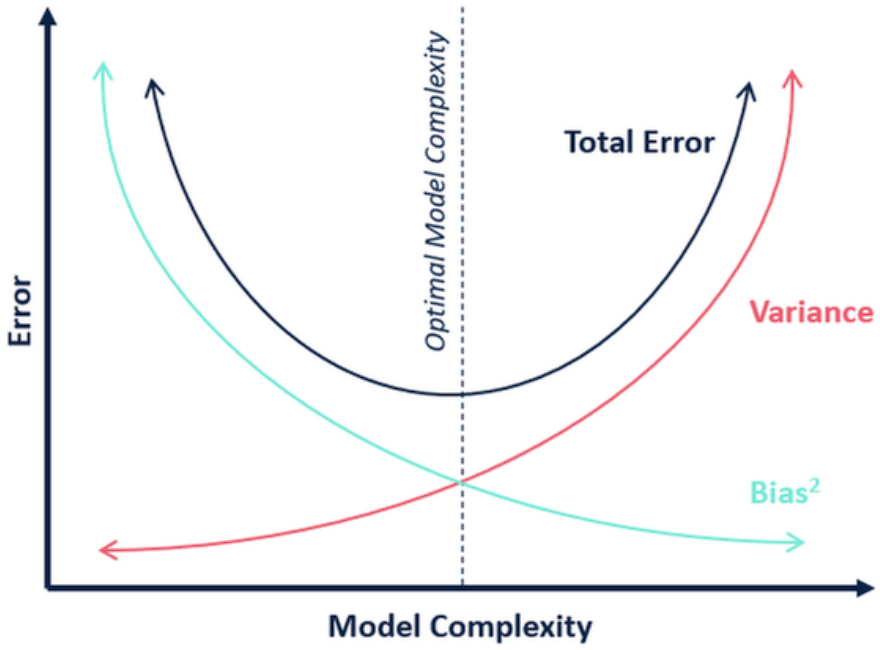
ASSIGNMENT 5

Dataset	E_{train}	E_{test}
MONK-1	0.0	0.1713
MONK-2	0.0	0.3079
MONK-3	0.0	0.0556

These results confirm that:
- MONK-2 is the most difficult dataset to learn with decision tree.
- MONK-1 is a bit hard to train because it involves a relationship between two attributes (a1 = a2) and we have trained the tree using one attribute at a time.
- MONK-3 has 5% of misclassification in the training set, so the performance is not perfect (but very good).

ASSIGNMENT 6

- When the tree is fully grown, we have the highest complexity for the model that leads to smallest possible bias but high variance.
- Pruning allows us to reduce complexity of the model, avoiding overfitting, and a better bias-variance trade off that leads to the minimum total error.



ASSIGNMENT 7

- V needs to be large enough to provide statistically meaningful instances.
- At the same time we need to have enough data to train the model effectively.