

Report 1: Dectrees

Group XXX: XXX, Riccardo Fragale

January 24, 2026

1 Assignment 0

Each one of the datasets has properties which makes them hard to learn. Motivate which of the three problems is most difficult for a decision tree algorithm to learn.

The bigger challenge for all the three datasets is the number of training samples compared to test samples. Let's consider the dataset **MONK1** as example to justify that: it has exactly 431 test samples while the number of training samples is 123. Considering that we don't have a validation set and we don't use a k-fold cross-validation, we have a really limited amount of data to train the model on. This reasoning is valid also for **MONK2** and **MONK3**. This makes the decision-tree less able to generalise and have a complete picture of the classification problem to be solved.