# 1

The true concepts underlying each MONK’s problem are given by table 1. Each one of the datasets has properties which makes them hard to learn. Can you guess which of the three problems is most difficult for a decision tree algorithm to learn?

# 2

# 3

# 4

# 5

Split the monk1 data into subsets according to the selected attribute using the function select (again, defined in dtree.py) and compute the information gains for the nodes on the next level of the tree. Which attributes should be tested for these nodes?

* There will be only one new node, in which we will test attribute A1 with gain 0.058. The gain for node 2 can be seen in the table below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | A4 | A5 | A6 |
| Monk1 | 0.0583 | 0.0052 | 0.0025 | 0.0424 | 0.0011 | 0.0012 |

**Assignment 3:**

|  |  |  |
| --- | --- | --- |
|  | **C\_train (% correctly classified)** | **C\_test (% correctly classified)** |
| **MONK-1** | **100%** | **83%** |
| **MONK-2** | **100%** | **69%** |
| **MONK-3** | **100%** | **94%** |

# 6