Monk test

October 18, 2021

1 The Monk's Problem

MONK's problem was the subject of the first international summit for the comparison of learning algorithms. The results of the study have been summarised in "The MONK's Problems - A Performance Comparison of Different Learning Algorithms" by S.B. Thrun et al.

One significant characteristic of this comparison is that it was performed by a collection of researchers, each of whom was an advocate of the technique they tested (often they were the creators of the various methods). In this sense, the results are less biased than in comparisons performed by a single person advocating a specific learning method, and more accurately reflect the generalization behavior of the learning techniques as applied by knowledgeable users.

There are three MONK's problems. The domains for all MONK's problems are the same (described below). One of the MONK's problems has noise added. For each problem, the domain has been partitioned into a train and test set.

Attribute Information:

- class: 0, 1
- a1: 1, 2, 3
- a2: 1, 2, 3
- a3: 1, 2
- a4: 1, 2, 3
- a5: 1, 2, 3, 4
- a6: 1, 2
- Id: (A unique symbol for each instance)

If we do not consider the class attribute, the input attributes are 7. We transformed them with a one-hot-encoding on all three Monks datasets, resulting in 17 binary inputs per pattern.

Number of Patterns: 432

Missing attribute values: None
Target concepts for each of the MONK problems:

- Problem MONK-1: (a1 = a2) or (a5 = red). From 432 possible examples, 124 were randomly selected for the training set. There were no misclassications.
- Problem MONK-2: EXACTLY TWO of a1 = 1, a2 = 1, a3 = 1, a4 = 1, a5 = 1, a6 = 1. From 432 possible examples, 169 were randomly selected. Again, there was no noise.
- Problem MONK-3: (a5 = 3 and a4 = 1) or (a5 != blue and a2 != octagon). From 432 possible examples, 122 were randomly selected. 5% misclassification.