Practical assignment: Reinforcement Learning

The assignment is carried out individually or in groups of maximally 2 persons.

The underlying Markov Decision Process of this assignment is identical to the one of the first assignment.

You may use any programming language to do this assignment.

# The assignment

Solve a Tower of Hanoi problem with three pins 1, 2 and 3 and two disks A and B. Disk A is larger than B. The goal is to move the two disks to pin 3 such that the lager disk A is at the bottom and the smaller disk B is at the top. Reaching this goal has a reward of: 100. The agent can move only one disk at the time. We do not forbid to put the larger disk A on top of the smaller disk B, but there is a penalty for doing this; i.e., a reward of: -10. To encourage the agent to solve the problem in a minimal number of steps, all actions that do not result in the goal state or the penalized state have a reward of: -1.

Our agent can make mistakes. When moving a disk from pin *i* to pin *j*, the agent may actually put the disk on pin *k* where *k*≠*i* and *k*≠*j*. The probability of making a mistake is: 0.1.

Find the optimal policy for the MDP using **Reinforcement Learning**, assuming that the agent does not know the above specified transition function and reward function. You may ignore the *exploitation* of the results that have been learned.

After reaching the absorbing state, continue the learning process in a randomly selected other state.

### Warnings

Note that the max-operator only applies to the term directly following the operator.

Note that the Q-value can never be higher than 100. Do you know why?

# Report

A small report must be handed in describing

* the states and the actions,
* a discussion of how you can learn the optimal result for every initial state,
* the optimal policy,
* the q-values of each state-action pair given the optimal policy,
* an evaluation of the convergence speed (as a function of the accuracy),
* the code of the implementation. If you use Java, hand in the source code and an executable jar-file.

Does the result differ from the result of the first assignment?

Mention in the header of the report your name and student number.