Homework 8

2-AIN-150, Winter 2018

Deadline: 8.1.2021, 23:59, boza@fmph.uniba.sk, email subject: Machine learning HW 8

Before you start solving the homework, please read the general instruction at the end of the document. Submitted solutions should be your own. Do not copy and do not try to find solution in literature or over the internet.

Reinforcement learning

Consider an empty road (a line). There is a traffic light at position 0. Your car is at position a (it might be negative) and currently it has speed 0. Your want to reach position 23, stop there and do not cross the red light.

We will use reinforcement learning to solve this task. The state will have three parts:

- Car position – integer from range $-50, 50$
- Rýchlosť auta –integer from range $-5, 5$ (negative speed means going backwards).
- Light color at traffic light and its duration – number from set $\{-5, -4, -3, -2, -1, 1, 2, 3, 4, 5\}$. Positive number f means, that green light will be there for $f - 1$ steps and negative number $-f$ means, that red will be there for $f - 1$ steps.

Set of states consists from all triplets from numbers above. State $(23, 0, x)$ is a good goal state (no actions from there) and states $(50, x, y)$ a $0(-50, x, y)$ are bad goal states (no actions from there).

Actions are from set $A = \{-1, 0, 1\}$. Action first changes speed of the car (by the same number as action has) and then we change position of the car (using the speed).

Transition probability is defined as: With probability 90% your car does what you want and with probability 5% it does one of other actions.

Reward is defined as: In good target state your reward is 100. In bad target state the reward is -100. In all other states your reward is -0.1.

Using your selected method (value iteration, policy iteration, q-learning) write a functions, which calculated the best policy.

Bonus points:

- 3 points – write Q-learning and value or policy iteration and compare speed of convergence.
- 10 bodov – write approximation of Q-function using suitable metho (linear regression, neural network) and asses its accuracy.

Usual rules apply: Do not use libraries, which simplify your work to oneliner.

Pokyny pre Python There is a file `template.py` in the package. You should fill out the function `get_policy()`. Your function should return the best policy. You are given various helper functions for getting next state, simulation of your solution, ... Program can be runned using `python template.py`.

General instructions

You should submit your code via moodle. **Add some short commentary about your methods into source code.**

Ideally submit your homeworks in Python (fill out required functionality from assignment). You can use different language if you really want, but you need also to add auxiliary functionality like reading input and output. But your solution should be runnable under Linux using open source software.