Prof. Joschka Boedecker, Gabriel Kalweit, Maria Huegle, Andreas Saelinger

Reinforcement Learning
Exercise 5

1 Q-learning: Pen and Paper

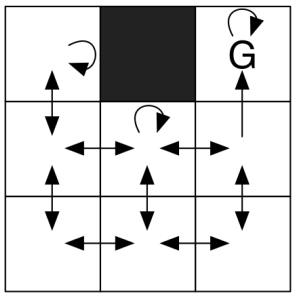

Figure 1: Grid MDP

Consider the deterministic MDP in Figure 1. There exists a terminal state G and a wall that cannot be entered. The agent remains in its current position if it chooses an action that moves it against the wall or off the grid. All transitions have a reward of -1. We discount with 0.5.

(a) How is the Q-learning update defined for a transition from state i to j, if j is a terminal state?

(b) We initialize all Q-values with 0. The agent starts in the upper left corner. It then moves one cell down, then one cell to the right and tries unsuccessfully to move one cell upwards (i.e. remains in its current cell), then moves one cell to the right and finally moves upwards into the terminal state. Which values of the Q-function change during this episode if we apply Q-learning with a learning rate of 1.0? Calculate the updated Q-function after this first episode. Repeat the calculation for a second identical episode.

(c) Calculate the optimal Q-values $Q_*(s, a)$ for all state-action pairs.

2 Q-learning: Implementation

This task are based on the Cliff Walking example from the book[1]. We already prepared a complete implementation of the environment that can be found in `cliff_walking.py`. All that is left to do for you is to implement the Q-learning algorithm:

```
q_learning(env, num_episodes, discount_factor=1.0, alpha=0.5, epsilon=0.1),
```

in `q_learning.py`.

Furthermore we provide tests in `exercise-05_test.py` as well as a visualization in `visualization.py`.

3 Experiences

Make a post in thread *Week 05: Temporal-Difference Learning* in the forum[2], where you provide a brief summary of your experience with this exercise and the corresponding lecture.

[1] `http://incompleteideas.net/book/RLbook2018.pdf#page=154`
[2] `https://ilias.uni-freiburg.de/goto.php?target=frm_1837317&client_id=unifreiburg`