Prof. Joschka Boedecker, Yuan Zhang, Jasper Hoffmann

Reinforcement Learning
Exercise 6

1 Dyna-Q

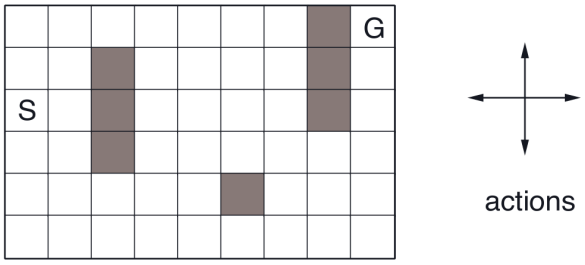

Figure 1: Grid MDP

Consider the deterministic MDP in Figure 1 from the book[1]. There exists a terminal state G and walls 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 0 except for those leading into the goal with a reward of $+1$. We discount with 0.95. You can find an implementation of the environment in `gridworld.py`.

Implement the Dyna-Q algorithm from the lecture in `dyna_q_learning.py`. Play around with the number of planning steps n. You can reuse your Q-learning implementation from last week.

2 Experiences

Make a post in thread *Week 06: Planning and 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=186`
[2] `https://ilias.uni-freiburg.de/goto.php?target=frm_2879356&client_id=unifreiburg`