

# Deep Learning Lab course: Exercise 4

Maria Huegle, University of Freiburg

December 1, 2017

## 1 Q-Learning

Consider the grid-world depicted in Figure 1. This grid-world has deterministic state transitions and an absorbing goal state G. The actions are moving up, down, left or right. The actions applicable in a specific state are depicted via the black arrows. Furthermore, the MDP features a wall a state that can not be accessed by the agents. Choosing an action that would lead into this state leave the agent where it is. All transitions have immediate reward of -1, only transitions within the goal-state are free (reward 0). Furthermore, we use a large discounting factor of  $\gamma = 0.5$ .

1. Write down the update-rule of Q-Learning for updating the Q-function after a transition from a state  $i$  to a state  $j$  using action  $u$  and observing immediate reward  $r(i, u)$ . How would you handle transitions to or within the goal state (which is absorbing, i.e. the agent can never transition out of it)?
2. Starting with a zero-initialized Q-function, the agent starts in the upper left corner, moves a cell down, one cell to the right, tries to move upwards, fails and ends in the same cell, moves a cell right and finally moves a cell upwards into the goal state, ending this episode. Determine, which Q-values would have been changed during this episode when using Q-learning with a learning rate of 1.0. Specify the improved Q-function after this initial episode.

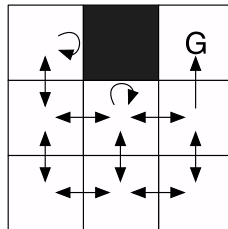


Figure 1: Grid-world

## 2 Deep Q-learning

Implement the deep q-learning agent for the simple maze task as described in the course repository [https://github.com/aisrobots/dl\\_lab\\_2017/tree/master/exercise4robotics](https://github.com/aisrobots/dl_lab_2017/tree/master/exercise4robotics). Hand in the exercise on 22.01 including a report of 2-3 pages. Work in groups of two.