Reinforcement Learning
Exercise 3

Jim Mainprice, Philipp Kratzer
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

May 5, 2020

Submission Instructions:

The submission deadline for this exercise sheet is 12.05., 23:55.

Put your answers into a single pdf. Your python code should be a single python script. Upload both files to ilias. Make sure that the code runs with *python3 yourscript.py* without any errors.

Group submissions of up to three students are allowed.

1 Proofs (5P)

a) Show that the Bellman optimality operator $\mathcal{T}$ is a γ-contraction. (2P)

$$(\mathcal{T}v)(s) = \max_a \sum_{s',r} p(s', r|s, a)[r + \gamma v(s')] \tag{1}$$

b) Asuming a general finite MDP (S, A, R, p, γ) where rewards are bounded: $r \in [r_{\min}, r_{\max}]$ for all $r \in R$. Prove the following equations. (3P)

$$\frac{r_{\min}}{1-\gamma} \leq v(s) \leq \frac{r_{\max}}{1-\gamma} \tag{2}$$

$$|v(s) - v(s')| \leq \frac{r_{\max} - r_{\min}}{1-\gamma} \tag{3}$$

2 Value Iteration (5P)

As in the previous exercise sheet, we will use the FrozenLake environment from gym (`https://gym.openai.com/envs/FrozenLake-v0/`). The code template can be found on github (`https://github.com/humans-to-robots-motion/rl-course`) in *ex03-dynp/ex03-dynp.py*.

a) Implement the value iteration algorithm (see lecture 3 slide 25) in the function *value_iteration*. Use the values for γ and θ that are given in the code. Initialize the value function $V(s)$ to 0 for all states. How many steps does it need to converge? What is the optimal value function? Put your answers into your submission pdf. (3P)

b) Compute the optimal policy from the value function and put it into your submission pdf. (2P)