Reinforcement Learning
Exercise 5

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

May 20, 2020

Submission Instructions:

The submission deadline for this exercise sheet is 26.05., 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 Random Walk (2P)

Recall the Random walk example presented in the lecture. From the results shown in the left graph (estimated value) it appears that the first episode results in a change in only V(A). What does this tell you about what happened on the first episode? Why was only the estimate for this one state changed? By exactly how much was it changed (assuming $\alpha = 0.1$)?

The episode ended on the far left end-state. V(A) reduced 0.05 at the end of the episode (considering gamma=1)

2 Sarsa and Q-learning on the FrozenLake (8P)

The code template can be found on github (`https://github.com/humans-to-robots-motion/rl-course`) in *ex05-td/ex05-td.py*.

a) Implement Sarsa and obtain and plot the state-value function, action-value function, and policy for the FrozenLake environment. Plot the average episode length as training continues. Put your plots into your submission pdf. (3P)

b) Implement Q-learning and obtain and plot the optimal state-value function, action-value function, and policy for FrozenLake. Put your plots into your submission pdf. What can you say about on-line performance during training in comparison to the performance of the optimal policy? (3P)

c) Explore how your results for a) and b) change if you switch to the non-slippery version (i.e. deterministic environment). Put your plots into your submission pdf. (1P)

d) Rerun your code for the larger FrozenLake environment. Put your plots into your submission pdf. (1P)