Reinforcement Learning
Exercise 4

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

May 11, 2020

Submission Instructions:

The submission deadline for this exercise sheet is 19.05., 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 Monte Carlo Methods vs Dynamic Programming (3P)

a) What are advantages of Monte Carlo methods over dynamic programming? Mention at least two. (2P)

b) Give an example environment where you would use a Monte Carlo method to learn the value function rather than using dynamic programming. Explain why. (1P)

2 Monte Carlo ES for blackjack (6P)

In this exercise we use the blackjack environment from gym (`https://github.com/openai/gym/blob/master/gym/envs/toy_text/blackjack.py`). The code template can be found on github (`https://github.com/humans-to-robots-motion/rl-course`) in *ex04-mc/ex04-mc.py*.

a) Consider the version of blackjack introduced in the lecture (Example 5.1 from Sutton and Barto). Implement first-visit Monte Carlo prediction (slide 8) for the given policy: stick if sum ≥ 20, else hit. Reproduce the figures on slide 11 and include them in your submission pdf. (3P)

b) Implement Monte Carlo ES and obtain the optimal policy and state-value function for blackjack. Put the optimal policy into your submission pdf (e.g. as 2 tables, one with usable ace and one without usable ace). (3P)