Reinforcement Learning
Prof. S. Harmeling

Monday, 11. May 2020
DUE 23:55h Monday, 18. May 2020

Exercise set #4

Solution should be submitted in teams of two if possible. Due to the current COVID-19 pandemic please submit your solution online using the sciebo file-drop folder. The link will be available in ILIAS. Please submit a single zip file with the following naming scheme: `username1-username2.zip` (e.g. `jadoe101-jodoe108.zip`). Allowed file extensions (of files within the zip file) are: `.pdf`, `.txt`, `.py` and `.ipynb`. Make sure the total file size does not exceed 10 MB.

1. **Monte Carlo learning (from Sutton and Barto [1])**

   (a) **Exercise 5.3:** What is the backup diagram for Monte Carlo estimation of q_π?

   5 points

   (b) **Exercise 5.10:** In the lecture we have discussed weighted importance sampling for off-policy Monte Carlo learning. Read up on the incremental version on page 109 of Sutton and Barto [1].

   Derive the weighted-average update rule (5.8) from the value function estimate (5.7):

$$V_n = \frac{\sum_{k=1}^{n-1} W_k G_k}{\sum_{k=1}^{n-1} W_k} \tag{5.7}$$

$$V_{n+1} = V_n + \frac{W_n}{C_n}\left[G_n - V_n\right], \tag{5.8}$$

   where $C_n = C_{n-1} + W_n$ and $C_0 = 0$.

   15 points

2. **Monte Carlo control (programming task, from Shimon Whiteson [2])**
   Implement Monte Carlo control for OpenAI's Blackjack environment.
   Follow the instructions in `exercises04.ipynb`.

   80 points

References

[1] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction.* MIT press, 2018.

[2] Shimon Whiteson. Introduction to reinforcement learning. `https://github.com/mlss-skoltech/tutorials_week2`, 2019.