

DRL Homework 02

Leon Schmid

April 2023

How-To homework

Please read carefully:

- Please use the [Homework submission form](#) to hand in your assignment. (Submissions via Email are also accepted, but generally cause much higher workload for the teaching staff)
- Deadline: Submission by Sat. 6.5. 11.59pm
- Before you submit the homework, make sure to have formed a group of three students and have signed into a respective group on StudIP
- Every single member of the group has to submit the homework via this form (i.e. three group members create 3 separate submissions via the form!)
- You are allowed and even encouraged to collaborate with fellow students in this class on this and any subsequent homework assignment.
- It's recommended to use either IPython Notebooks, or plain python files for code, with a Markdown (.md) file for Task 1

Contents

1 Homework Review	1
2 Learning a policy via MC - Policy Iteration	2

1 Homework Review

This task asks you to review two other groups' homework. The goal includes (1) for you to get a better understanding of contents by reviewing other groups submissions, (2) helping them understand how they could improve with their code (and possibly RL), and (3) help you improve by receiving valuable feedback from other groups. Step-by-step:

1. Coordinate with two other groups for mutual feedback. You may use the forum to achieve this, but we also try to match groups spontaneously at each QnA.

2. Take 15-30 min each to review their respective submissions. Write bullet points on your findings (both what your group should learn from their submission, and what the other group should improve)
3. Get together and discuss this feedback with representatives of all three groups in one of either the in-person or digital QnA sessions. Have one of the attending tutors as a 'referee' for any upcoming discussion and questions, and make sure they write down having refereed your group.
4. Denote the groups and respective tutor in the homework submission form

2 Learning a policy via MC - Policy Iteration

For the following work with last weeks implementation of your own gridworld! You may revise/change pieces of it, or ask other groups for access to their implementation of course.

- Implement tabular MC-estimate Policy Iteration
- Measure average Return-per-Episode and plot it against (1) episodes sampled, and (2) wallclock-time

For an outstanding submission:

- Visualize the State-Action Values in your gridworld during training at regular intervals, and provide a visualization of them (e.g. a series of images, best combine them into a short video clip)