FAKULTÄT FÜR WIRTSCHAFTSINFORMATIK
UND WIRTSCHAFTSMATHEMATIK

UNIVERSITÄT MANNHEIM

Prof. Dr. Leif Döring
Daniel Schmidt, Benedikt Wille

Reinforcement Learning

9. Exercise Sheet

27.04.2026

For this assignment, knowledge from lectures 1 to 20 is assumed.

1. Convergence of Q-Learning

The assumptions and definitions of Theorem 3.5.3 (Convergence of Q-Learning) are given. Moreover, let

$$F(Q)(s,a) \coloneqq \mathbb{E}_s^{\pi^a}\left[R(s,a) + \gamma \max_{a' \in \mathcal{A}_{S_1}} Q(S_1, a')\right]$$

and

$$\varepsilon_n(s,a) \coloneqq R(s,a) + \gamma \max_{a' \in \mathcal{A}_{s'}} Q_n(s', a') - F(Q_n)(s,a)$$

for all $(s,a) \in \mathcal{S} \times \mathcal{A}$ and $n \in \mathbb{N}$. Show that the sequence

$$Q_{n+1}(s,a) \coloneqq Q_n(s,a) + \alpha_n(s,a)\big(F(Q_n)(s,a) - Q_n(s,a) + \varepsilon_n(s,a)\big), n \in \mathbb{N}$$

almost surely converges to Q^*.

2. SARSA

Rewrite a k-armed Bandit as an MDP in such a way that SARSA (Algorithm 20 with ϵ_n-greedy policy updates and $\alpha(s,a) = \frac{1}{N(s,a)+1}$) corresponds to the ϵ_n-greedy algorithm introduced in Chapter 1. Check that both algorithms are equivalent.

3. *Programming task: Standard RL Libraries

The goal of this exercise is to get familiar with two key reinforcement learning libraries: Gymnasium and Stable Baselines3. This will prepare you to implement non-tabular algorithms in more complex environments later. Explore how the Stable Baselines3 Zoo workspace works by cloning it and playing around with small Gym environments. Additionally, you can use the chapter on the stable baselines repositories in our coding guidelines as learning resources. Focus on the following:

- How do you interact with Gymnasium environments? How are actions input, and how are states and rewards returned?
- What's the difference between OnPolicyAlgorithm and OffPolicyAlgorithm in Stable Baselines3? How does each work?

Don't worry too much about technical details like policy neural network architectures or replay buffers, although you are highly encouraged to research or ask about everything in many instances you do not need exact knowledge. Instead, aim for a solid high-level understanding of the key

components and how they fit together.

Hint: Take this exercise seriously, understanding these libraries and the repository structure will be essential when implementing algorithms on your own later!

The solution to the theoretical exercises will be discussed in the exercise class in B2 on May 18, 2026, at Mathelounge in B6 B301.