Reinforcement Learning
Prof. Milica Gašić

Release: Thursday, 12 December 2024
Discussion: Wednesday, 18 December 2024

Exercise set #10

You do not have to hand in your solutions to the exercises and they will **not** be graded. However, there will be four short tests during the semester. You need to reach ≥50% of the total points in order to be admitted to the final exam (Klausur). The tests are held at the start of a lecture (room 2522.U1.74) at the following dates:

Test 1: Thursday, 31 October 2024, 10:30-10:45
Test 2: Thursday, 21 November 2024, 10:30-10:45
Test 3: Thursday, 5 December 2024, 10:30-10:45
Test 4: Thursday, 9 January 2025, 10:30-10:45

Please ask questions in the RocketChat
The exercises are discussed every Wednesday, 14:30-16:00 in room 2512.02.33.

1. **Score function**

   Let $\pi_\omega(a \mid s)$ be a softmax policy defined as

   $$\pi_\omega(a \mid s) = \frac{\exp\left(x(s,a)^\top \omega\right)}{\sum_{a'} \exp\left(x(s,a')^\top \omega\right)}$$

   with weight vector $\omega \in \mathbb{R}^d$ and feature vector $x(s,a) \in \mathbb{R}^d$.

   Show that the score function of $\pi_\omega(a \mid s)$ is

   $$\nabla_\omega \log \pi_\omega(a \mid s) = x(s,a) - \sum_{a'} \pi_\omega(a' \mid s) x(s,a').$$

2. **REINFORCE**

   Implement REINFORCE with a softmax policy and apply it to the CartPole environment. Follow the instructions in the Jupyter notebook `reinforce.ipynb`.

Exercises by Stefan Harmeling, used with permission