

Exercise set #4

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 $\geq 40\%$ 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 25.22.U1.74) at the following dates:

Test 1: Thursday, 6 November 2025, 10:30-10:45

Test 2: Thursday, 27 November 2025, 10:30-10:45

Test 3: Thursday, 11 December 2025, 10:30-10:45

Test 4: Thursday, 15 January 2026, 10:30-10:45

Please ask questions in the RocketChat

The exercises are discussed every Wednesday, 14:30-16:00 in room 25.12.02.33.

1. Monte Carlo prediction

- (a) In Monte Carlo prediction the value function is estimated with the average return of collected episodes. Given an observed state s_t and return g_t , there are multiple ways to calculate the new value estimate:

(i) $N_{k+1}(s_t) = N_k(s_t) + 1$

$$G_{k+1}(s_t) = G_k(s_t) + g_t$$

$$V_{k+1}(s_t) = G_{k+1}(s_t) / N_{k+1}(s_t)$$

$$\text{with } N_0(s_t) = 0, G_0(s_t) = 0, V_0(s_t) = 0$$

(ii) $N_{k+1}(s_t) = N_k(s_t) + 1$

$$W_{k+1}(s_t) = W_k(s_t) + \frac{1}{N_{k+1}(s_t)}(g_t - W_k(s_t))$$

$$\text{with } N_0(s_t) = 0, W_0(s_t) = 0$$

Show that both approaches are equivalent, i.e., show by induction for all $k \geq 0$

$$V_k(s_t) = W_k(s_t).$$

- (b) Implement Monte Carlo prediction for tic-tac-toe with random moves. Follow the instructions in the Jupyter notebook `monte-carlo-prediction.ipynb`. What is the probability that the first player wins? Which initial action has the highest chance of winning?

2. Monte Carlo control

Implement Monte Carlo control and apply it to the Maze environment from the lecture. Follow the instructions in the Jupyter notebook `monte-carlo-control.ipynb`.