

Exercise: Q-Learning (ca. 2h)

In Q-Learning a value $Q(s,a)$ is learned for every state-action-combination (s,a) which describes how useful it is to do an action a in state s regarding the goal to maximize the expected rewards.

In the lecture we derived and talked about the update formula for these Q-values. Now use your 2D robot simulator and choose some learning scenario. You are free what has to be learned by your robot, how the world looks like and which sensors your robot has.

Implement the Q-value learning algorithm and observe what happens if you run the simulation for a longer time.

If you present your solution to your fellow students, first explain your scenario, then how you implemented the Q-value learning algorithm, then show your robot in its first learning steps and how the robot acts after a longer period of learning. Perhaps it makes sense that you prepare/record a video of the robot after a longer period of learning (e.g., using HyperCam).

Alternatively, you are able to restore saved Q-values that were learned before.