

REINFORCEMENT LEARNING  
Exercise 4  
Submit until **Thursday, December 12 at 2:00pm**



This week, we use the methods from last week for control. Please push your solutions to subdirectory `exercise-04` in your assigned git-repository. We are going to submit a `feedback.txt` in that directory. You find tests for the following tasks at `YOUR_REPO/exercise-04/tests/exercise-04_test.py`. Additionally, you find a visualization script `YOUR_REPO/exercise-04/scripts/visualization.py`. You again need matplotlib. Run the tests and the visualization with the usual commands.

## Preliminaries

This exercise is based on Lecture 5<sup>1</sup> from David Silver's RL course<sup>2</sup>. Watch before the upcoming meeting on Friday, December 1.

## 1 Off-Policy MC Control with Importance Sampling (10p)

This task is again based on the Blackjack environment from the lecture and the last exercise. Implement First Visit Off-Policy MC Control with Importance Sampling,

```
mc_control_importance_sampling(env, num_episodes, behavior_policy,  
                               discount_factor=1.0),
```

in `YOUR_REPO/exercise-04/scripts/off_policy_mc.py`.

## 2 Q-learning (10p)

The tests for this task are based on the Cliff Walking example from the lecture. An implementation can be found in `lib.envs.cliff_walking`. Implement the Q-learning algorithm,

```
q_learning(env, num_episodes, discount_factor=1.0, alpha=0.5, epsilon=0.1),
```

in `YOUR_REPO/exercise-04/scripts/q_learning.py`.

---

<sup>1</sup>[https://youtu.be/0g4j2k\\_Ggc4](https://youtu.be/0g4j2k_Ggc4)

<sup>2</sup><http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html>

### 3 Bonus: Experiences (1p)

Submit an `experiences.txt`, where you provide a brief summary of your experience with this exercise, the corresponding lecture and the last meeting. As a minimum, say how much time you invested and if you had major problems – and if yes, where.

---

Please push your solutions to subdirectory `exercise-04` in your assigned git-repository by **Thursday, December 12 at 2:00pm. Solutions after that or via email will not be accepted.**