

Reinforcement Learning, Tutorial 01

Philipp Kratzer

Machine Learning and Robotics Lab



University of Stuttgart
Germany

April 29, 2020

Announcements

- ▶ Submissions of first exercise sheet have been graded
- ▶ Second exercise sheet is available
- ▶ There was a small bug in the code, make sure you use the updated version!

Announcements

- ▶ We had 97 submissions from 191 students. More as expected!
- ▶ I will look into all the exercises and check whether what you did is correct
- ▶ But: I do not have the time to check every detail you write, keep your answers short and highlight the end results

1a

We have **2 actions**: a greedy and a non-greedy action. What is the probability that the greedy action is selected?

- ▶ For $\epsilon = 0$: 1.0 (only greedy selection)
- ▶ For $\epsilon = 1$: 0.5 (only random selection)
- ▶ For $\epsilon = 0.5$: 0.75 (greedy and random equally likely)

1b

Consider a k -armed bandit problem with $k = 4$
initial estimates of $Q_1(a) = 0$

We observe:

$$A_1 = 1, R_1 = 1, A_2 = 2, R_2 = 1, A_3 = 2, R_3 = 2, A_4 = 2, R_4 = 2, A_5 = 3, R_5 = 0.$$

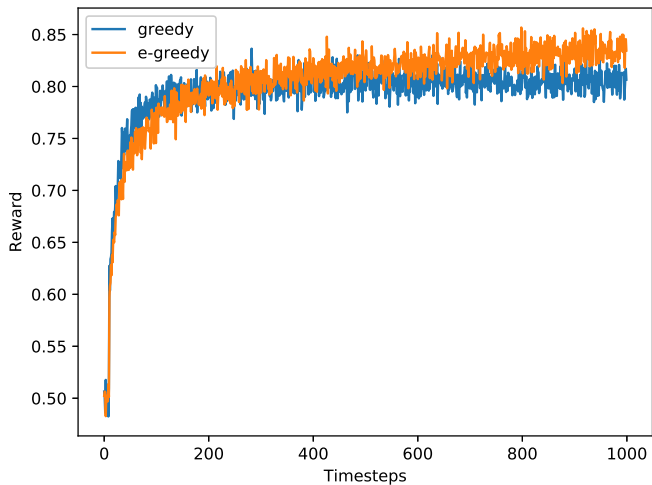
When definitely ϵ ?

- ▶ Step 2 and Step 5

When could be random?

- ▶ All time steps

2c



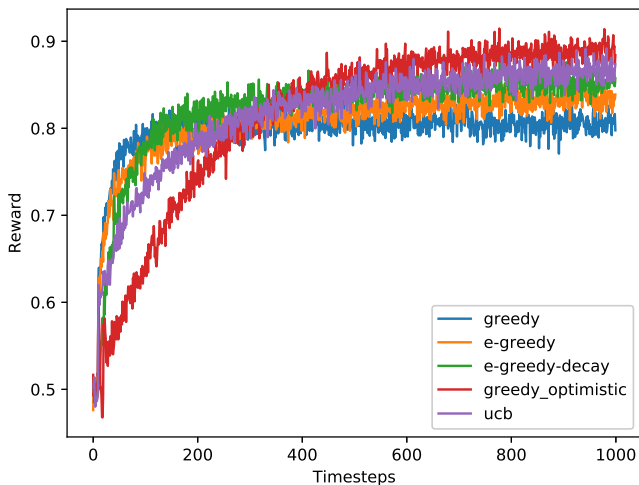
2d

A lot of possibilities:

- ▶ Decaying ϵ so that it converges to greedy
- ▶ Optimistic initialization could encourage exploration.
- ▶ One could give higher value to rarely visited states (e.g. UCB)
- ▶ Select very bad states less likely (e.g. softmax selection)
- ▶ If available, one could use prior information over the reward distributions (e.g. Thompson Sampling)

2d

Tradeoff: exploration vs exploitation



Next exercise sheet

- ▶ Exercise Sheet in Ilias
- ▶ Sourcecode on github
<https://github.com/humans-to-robots-motion/rl-course>
- ▶ There was a small bug so get latest version
- ▶ You need to install **openai gym** (<https://gym.openai.com/>)
- ▶ First environment FrozenLake (simple text-based gridworld)
<https://gym.openai.com/envs/FrozenLake-v0/>