

DATA.ML.200 Pattern Recognition and Machine Learning

Exercise Set 5: Reinforcement learning

1. `python` Q-learning. (5 points)

For this exercise you need to install the OpenAI Gym package (<https://www.openai.com/>). Instead of the FrozenLake environment used in the lecture demos, we use another similar environment, "Taxi". Launch Python and type the following commands:

```
$> python
>>> import gym
>>> env = gym.make("Taxi-v3")
>>> env.reset()
>>> env.render()
```

You should see a map and four pickup locations. Read the description of this environment from https://github.com/openai/gym/blob/master/gym/envs/toy_text/taxi.py. Use the actions encoded as 0-5 to solve the problem manually (render the updated map after each action):

```
>>> state, reward, done, info = env.step(1)
>>> env.render()
```

Your task is to implement Q-learning to solve the Taxi problem by finding a good (optimal) policy using the Q-table update rule. See the lecture notebook for an example.

During the Q-learning you should plot the current performance (mean \pm standard deviation) similar to the FrozenLake example. This should provide information how many episodes are needed and what is the best average total reward for $\gamma = 0.9$.

2. `python` Q-network. (10 points)

The goal of this exercise is to replace the Q-table with a neural network. However, you use the previous Q-table to training the network.

The network should have as many inputs as there are states in the Taxi environment (one-hot encoding). It also should have as many outputs as there are actions. The network is trained by sampling input-output pairs. Network structure, loss functions etc. you can decide yourself.

For input-output sampling (training data) you have two options:

- You sample the Q-table OR
- You use the Q-table to drive the Taxi, store input-output pairs and then train the network.

You may test the both training approaches to find out which one works better, and why? You also need to study how many training samples are needed until the Q-network encodes the same information as the Q-table.