

Reinforcement Learning

Exercise 11

Jim Mainprice, Philipp Kratzer, Yoojin Oh, Janik Hager

Machine Learning & Robotics lab, U Stuttgart

Universitätsstraße 38, 70569 Stuttgart, Germany

July 6, 2021

1 Maximum Entropy Inverse Reinforcement Learning on FrozenLake (10P)

The code template can be found on github (<https://github.com/humans-to-robots-motion/rl-course>) in *ex11-irl/ex11-irl.py*. For this exercise we will again use the FrozenLake environment. The code template includes a function to generate trajectories from an expert policy.

- a) Count state-action pair occurancies in the expert demonstrations and construct a policy from that. (2P)
- b) Use a linear reward representation and one-hot encoded features. A function for computing the optimal policy is already given in the template. Compute state frequencies $p(s|\psi)$ as given on slide 20. (3P)
- c) Implement the Maximum Entropy Inverse RL algorithm as described on slide 21. Plot the final reward function you obtained. (3P)
- d) What are advantages of c) over the naive approach in a)? Are there cases where c) works better than a) also in small tabular cases? Why or why not? (2P)