

Assignment 3

for Deep Reinforcement Learning WS2024

by

Kamil Bannasch

405231

Fynn Castor

540055

Eik Weishaar

507068

the 17. November 2024

1 Exercise 1.

2 Exercise 2.

2.1 Exercise 2a.

1. *Reflection Question:*

A random policy might be hindering in cases where a lot of different states with different actions exist. The randomness would make it likely that complete exploration of all state-value pairs would take a long time. Therefore a policy where actions are only chosen from the set of unexplored actions for the state, until all actions have been chosen in this state, would encourage exploration.

2. *Analysis:*

While state value estimation gives us a general grasp which states are desirable, it gives only little information about the impact of the specific action in each situation, especially if there is a random factor to the actions as in this example. Using the obvious action that should go to next state of highest value, would lead here in a lot of cases to the player falling in to the lake. As evident in the next exercise the optimal action is the one that avoids the possibility of falling into one of the holes.

2.2 Exercise 2b.

1. *Analysis:*

As evident in the last iteration the algorithm chooses actions, that make it least likely to end up unvoluntarily in a state from which falling into the lake is likely