UNIVERSITY OF FREIBURG

Exercise 2: Policy & Value Iteration

Nico Ott 4214197
Lior Fuks 4251285
Hendrik Vloet 4324249

November 23, 2017

1 Policy Iteration

- See according code file **policy_iteration.py**

2 Value Iteration

a)

- See according code file **value_iteration.py**

b)

- The policy iteration algorithm consists of two explicit components, the policy evaluation and the policy improvement step. These two steps are looped until convergence is achieved:
  $\pi_0 \xrightarrow{\text{eval}} v_{\pi_0} \xrightarrow{\text{improve}} \pi_1 \xrightarrow{\text{eval}} ... \xrightarrow{\text{improve}} \pi_* \xrightarrow{\text{eval}} v_*$
  Value iteration does not use an explicit policy evaluation step: $v_1 \rightarrow v_2 \rightarrow ... \rightarrow v_*$.
  After convergence to v_* the optimal policy π_* is instantly known. In praxis, convergence is obtained by check the difference of the two latest value functions. If it does only change for a small amount ($\sim$ 0.0001).
- One drawback of policy iteration is, that it can be computationally inefficient because we have to wait until it converges and the algorithm only does that in its limits (obviously). Value iteration converges much faster due to the lack of an explicit policy evaluation.
- value iteration uses the Bellman optimality equation in order to update their value function and policy iteration uses the bellman expectation equation and then greedily improves its policy.
- Similarity: value iteration is equivalent to policy iteration if policy iteration is terminated after one complete sweep of all states.

3 Experiences

- **Nico**
  - Invested Time:
    - Meeting: 0h; Was on my sisters wedding on Friday. Missed the meeting :-(
    - Lecture: 3h; Just watched the lecture and took notes (but way too late)
    - Exercise: 2h; I was sick from last tuesday on and had no energy for the assignment.
      Plus, there was my sisters wedding on Friday (and Saturday), where I had to organize some stuff as best man.
  - General:
    - It was really unfortunate for me, that nobody wrote a single thing in the Google Doc about last meeting.
    - I get the feeling that I'm one of the few that are actually using this opportunity.
    - Maybe the other students tend to use the Ilias forum if they have questions
    - But at least for flipped classroom some kind of protocol should always be available
- **Hendrik**
  - Invested time:
    - Meeting: 2h
    - Lecture 3: 3h
    - Exercise : $\sim$ 24h
  - General:
    - understanding issues with the unit test structure of the python environment
    - I had trouble comprehending why we had to go for a deterministic policy, thanks to rr114_uni-freiburg for helping out!
    - Sadly, protocol morale for the google doc seems rather low. I added some things regarding the video lecture and the according discussion last time.