Homework 1

Elias Dubbeldam, Ankur Satya

September 16, 2022

2.4 Coding Assignment - Dynamic Programming

1. See `codegra.de` for the notebook.

2. We have implemented the value iteration and policy iteration algorithms in the lab.

   Policy iteration cycles between the two steps of policy evaluation and policy improvement. In both steps, once has to loop over all states s. Each policy evaluation step stops once there is convergence.

   Value iteration cycles implicitly between two steps. The policy is evaluated only once for each s, where after the policy is updated directly. Hence, the policy evaluation step is similar as in policy iteration. However, it does not wait until the convergence before it updates (improves) the policy.

   - For a single iteration, the policy evaluation step has to converge for policy iteration before a new iteration can be started. This is not the case for value iteration, it directly continues after one update for each s. Therefore, we expect that a single value iteration is faster than a single policy iteration
   - Policy iteration takes fewer iterations in total. The policy is updated once the policy evaluation step has converged, which makes the policy improvement 'more effective' for an individual iteration. The policy update in value iteration happens 'less informed', because it happens directly after one update during the iteration, and does not wait until convergence.

2.5 Dynamic Programming

See the written notes below.