

Home Assignment 1

GROUP 27 (Alessia Knauss, Piergiuseppe Mallozzi, Tomasz Kosinski)

1.a

$$q_{\pi}(s, a) = \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_{\pi}(s')$$

1.b, 1.c, 1.e

See the code

1.d

$$v_{k+1}(s) = \max_{a \in \mathcal{A}} \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_k(s') \right)$$

1.f

We start from arbitrary initialized values functions for each state. We know the whole mdp, the probabilities, so all the dynamics. We use this knowledge to compute the rewards from each state-action-next_state and we apply Bellman optimality equation iteratively.

The more iterations we do the close we get to the optimal values of each state (\mathbf{v}^*).