CME 241: Assignment 4

Matteo Santamaria (msantama@stanford.edu)

January 29, 2021

1. Manual Value Iteration

We begin with the provided initial guess for V^*:

$$v_0(s_1) = 10.0, v_0(s_2) = 1.0, v_0(s_3) = 0.0$$

$$q_1(s_1, a_1) = \mathcal{R}_{s_1}^{a_1} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_1 s'}^{a_1} \cdot v_0(s') = 8 + (0.2)10 + (0.6)1 + (0.2)0 = 10.6$$

$$q_1(s_1, a_2) = \mathcal{R}_{s_1}^{a_2} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_1 s'}^{a_2} \cdot v_0(s') = 10 + (0.1)10 + (0.2)1 + (0.7)0 = 11.2$$

$$q_1(s_2, a_1) = \mathcal{R}_{s_2}^{a_1} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_2 s'}^{a_1} \cdot v_0(s') = 1 + (0.3)10 + (0.3)1 + (0.4)0 = 4.3$$

$$q_1(s_2, a_2) = \mathcal{R}_{s_2}^{a_2} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_2 s'}^{a_2} \cdot v_0(s') = -1 + (0.5)10 + (0.3)1 + (0.2)0 = 4.3$$

$$q_1(s_3, a_1) = 0$$

$$q_1(s_3, a_2) = 0$$

$$v_1(s_1) = \max_{a \in \mathcal{A}} \{q_1(s_1, a)\} = 11.2$$

$$v_1(s_2) = \max_{a \in \mathcal{A}} \{q_1(s_2, a)\} = 4.3$$

$$v_1(s_3) = 0$$

$$\pi_1(s_1) = a_2, \pi_1(s_2) = a_1$$

$$q_2(s_1, a_1) = \mathcal{R}_{s_1}^{a_1} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_1 s'}^{a_1} \cdot v_1(s') = 8 + (0.2)11.2 + (0.6)4.3 + (0.2)0 = 12.82$$

$$q_2(s_1, a_2) = \mathcal{R}_{s_1}^{a_2} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_1 s'}^{a_2} \cdot v_1(s') = 10 + (0.1)11.2 + (0.2)4.3 + (0.7)0 = 11.98$$

$$q_2(s_2, a_1) = \mathcal{R}_{s_2}^{a_1} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_2 s'}^{a_1} \cdot v_1(s') = 1 + (0.3)11.2 + (0.3)4.3 + (0.4)0 = 5.65$$

$$q_2(s_2, a_2) = \mathcal{R}_{s_2}^{a_2} + \gamma \cdot \sum_{s' \in \mathcal{S}} \mathcal{P}_{s_2 s'}^{a_2} \cdot v_1(s') = -1 + (0.5)11.2 + (0.3)4.3 + (0.2)0 = 5.89$$

$$q_2(s_3, a_1) = 0$$
$$q_2(s_3, a_2) = 0$$

$$v_2(s_1) = \max_{a \in \mathcal{A}} \{q_2(s_1, a)\} = 12.82$$

$$v_2(s_2) = \max_{a \in \mathcal{A}} \{q_2(s_2, a)\} = 5.89$$

$$v_2(s_3) = 0$$

$$\pi_2(s_1) = a_1, \pi_2(s_2) = a_2$$

To show that the relative ordering of the actions for each state will not change as $k \to \infty$, consider

$$q_k(s_1, a_1) - q_k(s_1, a_2) = 8 - 10 + (0.2 - 0.1)v_{k-1}(s_1) + (0.6 - 0.2)v_{k-1}(s_2)$$
$$q_k(s_1, a_1) - q_k(s_1, a_2) = -2 + (0.1)v_{k-1}(s_1) + (0.4)v_{k-1}(s_2)$$

Which will be greater than or equal to zero if $v_{k-1}(s_1) \geq 20 - 4v_{k-1}(s_2)$, which is satisfied when $k = 3$.

We can do a similar analysis and consider

$$q_k(s_2, a_2) - q_k(s_2, a_1) = -1 - 1 + (0.5 - 0.3)v_{k-1}(s_1)$$

Which is positive as long as $v_{k-1}(s_1) > 10$, again which is satisfied for $k = 3$.

Under these conditions, the optimal policy will not change in either state so the value function will not change. Therefore, for $n > 3$, the optimal policy will remain fixed.