Rienforcement learning: Assignment 2

Oluwatomilayo, Adegbite
500569283xercise 3.19

Nikolas, Maier
500461990

February 21, 2019

1

1.1 Question 4

$$R_s^a = \sum_{s'} P_{ss'}^a R_{ss'}^a \tag{1}$$

$$V^\pi(s) = R_s^{\pi(s)} + \gamma \sum_{s' \in S} P_{ss'}^{\pi(s)} V^\pi(s') \tag{2}$$

Let $\pi' = greedy(V_{apr})$

$$\pi'(s) = arg \max_{a \in A} \left\{ R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_{apr}(s') \right\} \tag{3}$$

$$L^\pi(s) = V^*(s) - V^{\pi'}(s) \tag{4}$$

$$\max_{s \in S} \left\{ L^{\pi'}(s) \right\} \leq \frac{2\gamma\epsilon}{1-\gamma} \tag{5}$$

Must show that if $\mid V^*(s) - V_{apr}(s) \mid \leq \epsilon$ then $\max_{s \in S} \left\{ L^{\pi'}(s) \right\} \leq \dfrac{2\gamma\epsilon}{1-\gamma}$ for every state s, that is the Loss between the optimal value function and the approximated value function is less then accuracy exptected the policy is optimal.

1.1.1 Proof

Since

$$V^\pi(s) = R_s^{\pi(s)} + \gamma \sum_{s' \in S} P_{ss'}^{\pi(s)} V^\pi(s') \tag{6}$$

$$\max_{s \in S} \left\{ L^{\pi'}(s) \right\} \leq \frac{2\gamma\epsilon}{1-\gamma} \tag{7}$$

$$\max_{s \in S} \left\{ V^*(s) - V^{\pi'}(s) \right\} \leq \frac{2\gamma\epsilon}{1-\gamma} \tag{8}$$

$$\leq \frac{2\gamma\epsilon}{1-\gamma} \tag{9}$$

$$R_s^{*(s)} + \gamma \sum_{s' \in S} P_{ss'}^{*(s)} V^*(s') - R_s^{\pi(s)} + \gamma \sum_{s' \in S} P_{ss'}^{\pi(s)} V^{\pi}(s') \leq \frac{2\gamma\epsilon}{1-\gamma} \tag{10}$$

$$\tag{11}$$