Rienforcement learning: Assignment 2

Oluwatomilayo, Adegbite 500569283

Nikolas, Maier 500461990

February 21, 2019

1

1.1 Exercise 3.17

Must give action value q_π(s,a) in terms of q_π(s',a')

$q_\pi(\text{s,a}) = E_\pi[G_t \mid S_t = \text{s} \ , \ A_t = \text{s} \ ]$

$= E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, A_t = s]$

$V_\pi(\text{s}) = \sum_a \pi(\text{s,a}) \sum P^a_{ss'} \ (R^a_{ss'} + \alpha \ V_\pi(\text{s'}))$

$q_\pi(\text{s,a}) = \sum_{s'} P^a_{ss'} \ (R^a_{ss'} + \sum_{a'} \ \alpha \ q_\pi(\text{s',a'}))$

To get the value of all possible successor states sum over s', multiply the probability of moving from current state to next state given action a by the result of the reward of that action and state transition and the sum of all possible actions in s' times a discount times the value of q_π(s',a')

1.2 Exercise 3.19

$q_\pi(\text{s,a}) = E[G_t \mid S_t = \text{s} \ , \ A_t = \text{s} \ ]$

$q_\pi(\text{s,a}) = E \ [R_{t+1} + \alpha \ V_\pi(\ s'_{t+1}) \mid S_t = \text{s} \ , \ A_t = \text{s} \ ]$