RL - HW 2

Shaurya Goel
2016194

Ans-1

s	a	s'	r	$P(s', r \mid s, a)$
high	search	high	r_{search}	α
high	search	low	r_{search}	$1-\alpha$
high	wait	high	r_{wait}	1
low	search	high	-3	$1-\beta$
low	search	low	r_{search}	β
low	wait	low	r_{wait}	1
low	recharge	high	0	1

If s = high, we can search or wait & get corresponding rewards. If s = low, we can search, wait or recharge. If we search, then, we can drain the battery & have to be rescued; we get -3 reward. If we choose to recharge, then we get 0 reward & transition to high state.

Ans-5

$$v_*(s) = \max_{a \in A(s)} q_*(s,a) \quad \forall s$$

Ans-3 3.15

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots$$

new rewards are —

$$R_{t+1} + C, \; R_{t+2} + C \; \ldots$$

$$G_t' = (R_{t+1} + C) + \gamma (R_{t+2} + C) + \gamma^2 (R_{t+3} + C) + \ldots$$

$$= (R_t + \gamma R_{t+2} + \gamma^2 R_{t+2} + \ldots) + C + \gamma C + \gamma^2 C + \ldots$$

$$\Rightarrow G_t' = G_t + \frac{C}{1-\gamma} \quad ; \; v_c = \frac{C}{1-\gamma}$$

As G_t' is changed by a constant term
$\Rightarrow$ value of all the states is increased by v_c.

As $v_\pi'(s) = E[G_t']$
$\Rightarrow v_\pi'(s) = v_\pi(s) + v_c$

3.16 Now, $G_t = R_{t+1} + \gamma R_{t+2} + \dots \gamma^{N-1} R_{t+N}$

Let N = episode length

$$G_t' = (R_{t+1} + c) + \gamma (R_{t+2} + c) \dots \gamma^{N-1} (R_{t+N} + c)$$

$$= (R_{t+1} + \gamma R_{t+2} + \dots \gamma^{N-1} R_{t+N}) + c + \gamma c + \dots \gamma^{N-1} c$$

$$\Rightarrow G_t' = G_t + c\left(\frac{1-\gamma^N}{1-\gamma}\right)$$

This would change the task, as G_t' depends on episode length N.

$$v_\pi(S_t) = E\left[R_{t+1} + \gamma v_\pi(S_{t+1})\right]$$

current value of state depends on future values of successor states. To both of these we add different constants according to N.

If N is small, value of all states change by less amount than for states with large N.