Name -
Ishaan Bassi
Roll no - 2016238

Assignment 2 Theory

Important → where $p(s', r|s, a) > 0$
i.e $p(s', r|s, a) = p(s|s, a)$

Q1. The table will stay the same except for the ~~first~~ rows where $p(s'|s, a) = 0$ are removed. This is due to the fact that only a single reward is possible with each triplet of (s, a, s'). Now we can see that for each (s, a), $\sum_{s', r} p(s', r|s, a) = 1$ as follows

~~high~~ S	a	
high	search	$(\alpha + 1 - \alpha) = 1$
low	search	$(1 - \beta + \beta) = 1$
high	wait	$(1 + 0) = 1$
low	wait	$(0 + 1) = 1$
low	recharge	$(1 + 0) = 1$

Q6. (BUG)

The following two changes are made in the code -

(1) The policy π is taken to be ~~either~~ stochastic instead of deterministic ie all actions ~~have~~ have some probability.

(2) If more than one action comes out to be optimal they are given equal probabilities.

Date / /

Page no.

(a)

Q 3. Exercise 3.15

No, the signs of the rewards do not matter, instead only their values relative to each other are important. This is due to the fact ~~or~~ that adding a constant to all the rewards does not change the relative distribution of state value function*. This can be seen as follows -

*Contd in the continuing case.

Continuing Case

$$V_\pi(s) = E\left[G_t \mid S_t = s\right]$$

$$= E\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots \mid S_t = s\right]$$

Adding a constant c to all rewards -

$$V'_\pi(s_t) = E\left[R_{t+1} + c + \gamma(R_{t+2} + c) \cdots \mid S_t = s\right]$$

$$= E\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots \mid S_t = s\right]$$

$$+ c + \gamma c + \gamma^2 c + \cdots$$

$$= E\left[R_{t+1} + \gamma R_{t+2} \cdots \mid S_t = s\right] + \frac{c}{1-\gamma}$$

$$= E\left[G_t\right]^{S_t = s} + \underbrace{\frac{c}{1-\gamma}}_{\downarrow V_c}$$

Date

Exercise 3.16

Page no.

Episodic Case

$$V_\pi(s) = E[G_t \mid s_t = s]$$

$$= E[R_{t+1} + c + \gamma(R_{t+2} + c) + \cdots + \gamma^{k-1}(R_{t+k} + c)]$$

$$= E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots] + c\left[\frac{1-\gamma^k}{1-\gamma}\right]$$

For smaller 'k', the term is smaller. Hence in this case, the states that are closer to the terminal state become less desirable now. Hence there is a relative change in state values.

Q5. $$V_*(s) = \max_a q^*(s,a) \quad \text{--- (1)}$$

$$\Rightarrow V_*(s) = \max_{a \in A(s)} E[G_t \mid s_t = s, A_t = a]$$

$$= \max_a \sum_{s'} \sum_r p[s', r \mid s, a] \cdot [r + \gamma V_*(s')]$$

using (1) again

$$v_*(s) = \max_a \sum_{s'} \sum_r p[s', r \mid s, a][r + \gamma \max_a q^*(s', a)]$$