Homework 2

Ans3. Ex 3.15 →

(a) No, sign of the rewards doesn't matter, only the intervals between them matters, This is because when we add a constant c to all the possible rewards, another constant V_c is added to all $v_\pi(s)$ $s \in S$. Thus relative Expected reward among states remain same. We can prove this in the following way using bust By Bellman equation → (b)

$$v_\pi(s) = \sum \pi(a|s) \sum_{s',r} p(s',r|s,a)\left(r + \gamma v_\pi(s')\right)$$

now we add constant c to r

$$\Rightarrow v_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left(r + c + \gamma v_\pi(s')\right)$$

$$\Rightarrow v_\pi(s) = \sum_a \pi(a|s)\left(\sum_{s'r} p(s',r|s,a)\left(r + \gamma v_\pi(s')\right) + \sum_{s'r} p(s'r|s,a) \times c\right)$$

$$= \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a) * \left(r + \gamma v_\pi(s')\right) + c$$

~~$\sum_a \pi(a|s) \times$~~ ($\because \sum_a \sum_{s'} \sum_r \pi(a|s)\, p(s',r|s,a) = 1$).

(b) $G_t = R_{t+1} + \gamma R_{t+12} \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$

adding constant c to reward.

$$G_t = \sum_{k=0}^{\infty} \gamma^k (R_{t+k+1} + c)$$

$$= \sum_{R=0}^{\infty} R_{t+k+1} + c \sum_{k=0}^{\infty} \gamma^k$$

$$G_t = \sum_{k=0}^{\infty} \gamma^k (R_{t+k+1}) + \frac{C}{1-\gamma}$$

$$V(s) = \mathbb{E}\left[G_t \mid S_t = s \right]$$

$$= \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k (R_{t+k+1}) + \frac{C}{1-\gamma} \mid S_t = s \right]$$

$$V(s) = \frac{C}{1-\gamma} + \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k (R_{t+k+1}) \mid S_t = s \right]$$

$$\therefore \boxed{V_C = \frac{C}{1-\gamma}}$$

We can clearly see that adding constant C to all the rewards adds $\frac{C}{1-\gamma}$ to $V(s)$. Hence proved.

Ex 13.16

g_t would have an effect in the case of episodic task as in that case if the length of the episode is T

then, $\sum_{k=0}^{T-1} C\gamma^T$ would be added, that is equal to $C\left(\frac{1-\gamma^T}{1-\gamma}\right)$

thus reward would change more as we progress towards the end of an episode.

Ans ~~$v_*(s) = \max_{a \in A(s)} q_{\pi_*}(s,a)$~~

Ans $v_*(s) = \max_{\pi} v_\pi(s)$.

$q_*(s,a) = \max_{\pi} q_\pi(s,a)$.

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

Q1

s	a	s'	r	p(s', r | s, a)
high	search	low	0	$p_1 = 1-\alpha-r_{search}$
"	"	"	1	$p_2 = r_{search}$
"	"	"	-3	0
"	"	high	0	$p_3 = \alpha - r_{search}$
"	"	"	1	$p_4 = r_{search}$
"	"	"	-3	0
low	"	low	0	$p_5 = \beta - r_{search}$
"	"	"	1	$p_6 = r_{search}$
"	"	"	-3	0
"	"	high	0	0
"	"	"	1	0
"	"	"	-3	$1-\beta$
high	wait	low	0	0
"	"	"	1	0
"	"	"	-3	0
"	"	high	0	$p_7 = 1-r_{wait}$
"	"	"	1	$p_{10} = r_{wait}$
"	"	"	-3	0
low	"	low	0	$p_{11} = 1-r_{wait}$
"	"	"	1	$p_{12} = r_{wait}$
"	"	"	-3	0
"	"	high	0	0
"	"	"	1	0
"	"	"	-3	0
low	recharge	high	0	1
"	"	"	1	0
"	"	"	-3	0
"	"	low	0	0
"	"	"	1	0
"	"	"	-3	0

Now, let's calculate p_i with the help of given information

$p_1 + p_2 = 1-\alpha$, $p_1 \times 0 + p_2 \times 1 = \gamma_{search}$

$\Rightarrow p_2 = \gamma_{search}$ $p_1 = 1-\alpha-\gamma_{search}$.

$p_3 + p_4 = \alpha$, $p_3 \times 0 + p_4 = \gamma_{search}$

$\Rightarrow p_4 = \gamma_{search}$, $p_3 = \alpha - \gamma_{search}$

$p_6 + p_6 = \beta$ $p_5 \times 0 + p_6 \times 1 = \gamma_{search}$

$\Rightarrow p_6 = \gamma_{search}$ $p_5 = \beta - \gamma_{search}$.

~~$p_7 + p_8 = 1-\beta$~~

$p_9 + p_{10} = 1$ $p_9 \times 0 + p_{10} = \gamma_{wait}$

$\Rightarrow p_{10} = \gamma_{wait}$ $p_9 = 1 - \gamma_{wait}$

$p_{11} + p_{12} = 1$ $p_{11} \times 0 + p_{12} = \gamma_{wait}$

$\Rightarrow p_{12} = \gamma_{wait}$ $p_{11} = 1 - \gamma_{wait}$