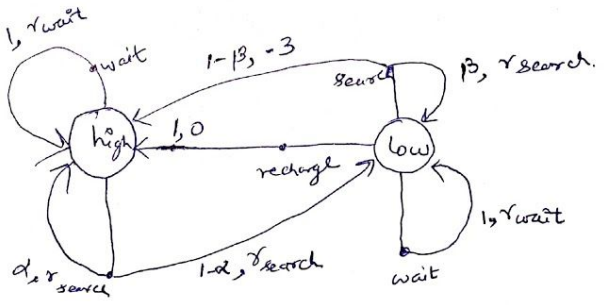
Assignment 2

Answer 1.

Table given below is.

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



Answer 3, part I

The value function of a state s under a policy π, denoted $v_\pi(s)$, is the expected return when starting in s and following π thereafter.

For MDPs v_π is defined as

$$v_\pi(s) = E_\pi\left[G_t \mid S_t = s \right]$$

$$= E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s \right] \text{ for}$$

all $s \in S$

Adding constant c to all the rewards

$$\hat{v}_\pi(s) = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k \hat{R}_{t+k+1} \mid S_t = s \right]$$

where $\hat{R}_{t+k+1} = R_{t+k+1} + c$

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

$$+ E_\pi\left[\sum_{k=0}^{\infty} \gamma^k c \mid S_t = s \right]$$

$$= V_\pi(s) + c\sum_{k=0}^{\infty} \gamma^k$$

$$= V_\pi(s) + \underbrace{c \times \frac{1}{1-\gamma}}_{\text{constant } V_c}$$

$$V_c = \frac{c}{1-\gamma}$$

Since V_c constant is added to the values of all states, and thus it does not affects the relative values of any states under any policies.

Answer 3 - part II

For episodic task. Let us assume task terminates after time step T.

$$\hat{v}_\pi(s) = v_\pi(s) + C\sum_{k=0}^{T} \gamma^k$$

$$= v_\pi(s) + C\left(\frac{1-\gamma^T}{1-\gamma}\right)$$

$$= v_\pi(s) + v_c$$

where $v_c = C\left(\frac{1-\gamma^T}{1-\gamma}\right)$

Since a constant is added to the values of all states, like previous case it does not affects the relative values of any state under any policies.

Answer 5.

The optimal value functions are recursively related by the Bellman optimality equation.

$$V_*(s) = \max_a q_*(s,a)$$

Deterministic case

If,

$V_*(s)$ ← optimal value function

and $q_*(s,a)$ ← optimal action value function.

then

$$q_*(s,a) = r(s,a) + \gamma V_*(s')$$

$r(s,a)$ ← transition reward

γ ← discounting factor.

s' ← next state.

Stochastic case

$$q_*(s,a) = r(s,a) + \gamma \sum_{s'} P(s'|s,a) V_*(s')$$

when $P(s'|s,a)$ is the transition probability to new state s'

So,

$$\boxed{V_*(s) = \max_a q_*(s,a)}$$