Name: Kartik . Bharadwaj

Roll No.: CS20S020

"I Pledge that I haven't copied or given any unauthorized assistance on this exam".

Q1) c

Q3) Let the discount factor be α for a disc. MDP.

We know the cum. disc. reward $R \in \mathbb{R}$ is:

$$R = \sum_{t=0}^{\infty} \alpha^t r(x_t).$$

(a)

$$J(x) = E[R \mid x_0 = x]$$

$$J(x) = E\left[\sum_{t=0}^{\infty} \alpha^t r(x_t) \mid x_0 = x\right]$$

$$J(x) = r(x) + E\left[\sum_{t=1}^{\infty} \alpha^t r(x_t) \mid x_0 = x\right]$$

$$J(x) = r(x) + \alpha E\left[E\left[\sum_{t=1}^{\infty} \alpha^{t-1} r(x_t) \mid x_0 = x, x_1 = x'\right]\right]$$

$$J(x) = r(x) + \alpha \sum_{x' \in X} P(x' \mid x) \cdot J(x').$$

$$M(x) = E\left[R^2 \mid x_0 = x\right]$$

$$\Rightarrow M(x) = E\left[\left(\sum_{t=0}^{\infty} \alpha^t r(x_t)\right)^2 \,\middle|\, x_0 = x\right]$$

$$\Rightarrow M(x) = E\left[\left(r(x) + \sum_{t=1}^{\infty} \alpha^t r(x_t)\right)^2 \,\middle|\, x_0 = x\right]$$

$$\Rightarrow M(x) = E\left[(r(x))^2 + 2r(x)\left[\sum_{t=1}^{\infty} \alpha^t r(x_t)\right] + \left[\sum_{t=1}^{\infty} \alpha^t r(x_t)\right]^2 \,\middle|\, x_0 = x\right]$$

$$\Rightarrow M(x) = r(x)^2 + 2r(x)\,E\left[\sum_{t=1}^{\infty} \alpha^t r(x_t) \,\middle|\, x_0 = x\right]$$

$$+ E\left[\left(\sum_{t=1}^{\infty} \alpha^t r(x_t)\right)^2 \,\middle|\, x_0 = x\right]$$

$$\Rightarrow M(x) = r(x)^2 + 2\alpha r(x) \sum_{x' \in X} P(x'|x) J(x')$$

$$+ \alpha^2 \sum_{x'} P(x'|x) M(x') \rightarrow \text{Using } J(x) \text{ from (a).}$$

$\boxed{\text{Part(b)}}$

We know variance $= E(X^2) - E(X)^2$

$$\Rightarrow V(x) = M(x) - J(x)^2$$

$$\Rightarrow V(x) = r(x)^2 + 2\alpha r(x) \sum_{x'} P(x'|x) J(x') + \alpha^2 \sum_{x'} P(x'|x) M(x')$$

$$- \Big[\underbrace{r(x) + \alpha \sum_{x'} P(x'|x) J(x')}_{J(x)}\Big]^2$$

$$\Rightarrow V(x) = r(x)^2 + 2\alpha r(x) \sum_{x'} P(x'|x) J(x')$$

$$+ \alpha^2 \left[\sum_{x'} P(x'|x) V(x') + \sum_{x'} P(x'|x) J(x')^2 \right] - J(x)^2$$

$$\Rightarrow \boxed{V(x) = r(x)^2 + 2\alpha r(x) \sum_{x'} P(x'|x) J(x') + \alpha^2 \sum_{x'} P(x'|x) J(x')^2 - J(x)^2 + \alpha^2 \sum_{x'} P(x'|x) V(x')}$$

We now have $V(x)$ in the following form,

$$\boxed{V(x) = \Psi(x) + \alpha^2 \sum_{x' \in X} P(x'|x) V(x')}$$

where $\Psi(x) = r(x)^2 + 2\alpha r(x) \sum_{x'} P(x'|x) J(x') + \alpha^2 \sum_{x'} P(x'|x) J(x')^2 - J(x)^2$

$$\Rightarrow \Psi(x) = [r(x) + J(x)][r(x) - J(x)] + 2\alpha r(x) \sum_{x'} P(x'|x) J(x') + \alpha^2 \sum_{x'} P(x'|x) J(x')^2$$

$$\Rightarrow \Psi(x) = [r(x) + J(x)]\left[-\alpha \sum_{x'} P(x'|x) \cdot J(x')\right]$$ → From a 3(a) eqn for $J(x)$

$$+ 2\alpha r(x) \sum_{x'} P(x'|x) J(x')$$

$$+ \alpha^2 \sum_{x'} P(x'|x) J(x')^2$$

$$\Rightarrow \Psi(x) = [r(x) - J(x)]\,\alpha \sum_{x'} P(x'|x) J(x') + \alpha^2 \sum_{x'} P(x'|x) J(x')$$

$$\Rightarrow \Psi(x) = -\alpha^2 \left[\sum_{x'} P(x'|x) J(x')\right]^2 + \alpha^2 \sum_{x'} P(x'|x) J(x')^2$$

$$\Rightarrow \boxed{\varphi(x) = \alpha^2 \left[\sum_{x'} P(x'|x) J(x')^2 - \left(\sum_{x'} P(x'|x) J(x') \right)^2 \right]}$$

Hence, proved.

Q4)

States $= \{G, B\}$.

$$\left.\begin{aligned} P(B|G) &= p_1 \\ P(G|B) &= p_2 \\ P(B|B) &= q_2 = 1 - p_2 \\ P(G|G) &= 1 - p_1 = q_1 \end{aligned}\right\} \text{Transition probabilities}$$

Single stage cost

$$g_G = -A \; ; \; g_B = B.$$

$$J^*(G) = -A + \gamma \left[P(B|G) J^*(B) + P(G|G) J^*(G) \right]$$

$$J^*(B) = B + \gamma \left[P(G|B) J^*(G) + P(B|B) J^*(B) \right]$$

$$\Rightarrow J^*(G) = -A + \gamma p_1 J^*(B) + \gamma q_1 J^*(G) \quad -(1)$$

$$J^*(B) = B + \gamma p_2 J^*(G) + \gamma q_2 J^*(B) \quad -(2)$$

from (2), we have:

$$J^*(B) = \frac{B + \gamma p_2 J^*(G)}{1 - \gamma q_2} \quad -(3)$$

Q4) continued

$$\Rightarrow J^*(B) = \frac{B + \dfrac{\gamma p_2(\gamma B p_1 - A(1-\gamma q_2))}{1-\gamma q_1 - \gamma q_2}}{1-\gamma q_2}$$

$$\Rightarrow \boxed{J^*(B) = \frac{B - \gamma B q_1 - \gamma B q_2 + \gamma^2 B p_1 p_2 - \gamma A p_2 + \gamma^2 A p_2 q_2}{(1-\gamma q_2)(1-\gamma q_1 - \gamma q_2)}} \rightarrow (5)$$

Q4

(b) $\therefore$

$$\min\{M + J^*(a), J^*(a)\}$$

$$\Rightarrow \min\{M, 1\} + J^*(a)$$

$\Rightarrow$ whenever cost $M < 1$, it is optimal to buy a new machine.

Q2) True since $\pi_3(i)$ is taking the optimal action w.r.t $\pi_1(i)$ & $\pi_2(i)$. ~~However, it is not necessary that~~

However, it is not necessary that $J_{\pi_3}(i) = \min\{J_{\pi_1}, J_{\pi_2}\}$

Since is π_3 is a better policy than π_1 & π_2,

$$J_{\pi_3} \leq \min\{J_{\pi_1}, J_{\pi_2}\}.$$

References for all my solution!

Q1) Internet (Google)

Q3) TD Methods for variance of reward to go (The authors prove for ~~the~~ an SSP. But I've converted it into a disc. MDP)