Divyam Anshumaan 2017147

RL · HW-2

Date:

Page No. 1

Q3

a) $G_t = R_{t+1} + \gamma R_{t+2} \cdots = \sum_{i=0}^{\infty} \gamma^i R_{t+1+i} = R_{t+1} + \gamma G_{t+1}$

now ~~att~~ adding reward c to all rewards, we have $V_\pi(s) = E_\pi(R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s)$

or $V'_\pi(s) = E_\pi(R_{t+1} + c + \gamma \cdot R_{t+1} + \gamma \cdot c + \cdots)$

$= E_\pi(c + \gamma c + \gamma^2 c + \cdots) + E_\pi(R_{t+1} + \gamma V_\pi(S_t$

But c is a constant, so $V'_\pi(s) =$

$c(1 + \gamma + \gamma^2 + \cdots) + E_\pi(R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s)$

$= c \cdot \frac{1}{1-\gamma} + V_\pi(s)$ (for $0 \leq \gamma < 1$)

$\therefore$ Gain for any state $s = V'_\pi(s) - V_\pi(s) = \left(\frac{c}{1-\gamma}\right)$

and $\boxed{v_c = c/1-\gamma}$

b) if $G_t = R_{t+1} + \gamma \cdot R_{t+2} + \cdots \gamma^{T-t-1} R_T$, then

$V_\pi(s) = E_\pi(R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t-1} R_T \mid S_t = s)$

or $V'_\pi(s) = E_\pi(R_{t+1} + c + c \cdot \gamma + \gamma \cdot R_{t+2} + \cdots + c^{T-t-1}\gamma^{T-t-1}$
$+ \gamma^{T-t-1} R_T \mid S_t = s)$

$\Rightarrow V'_\pi(s) = E_\pi(c(1 + \gamma + \cdots \gamma^{T-t-1})) + V_\pi(s)$

Scanned by CamScanner

$$= c \cdot \frac{1-\gamma^{T-t}}{1-\gamma} + v_\pi(s).$$

$\therefore$ Now the value gain is not the same as compared to ealier.

Q5. $v_*(s) = \max_{a \in A(s)} q_*(s,a)$