Lec-9, 1T567, 24-25

Without proof - there exists at least one policy that is better than or equal to all other policies. This is the optimal policy (or policies) $\triangleq \pi_*$

↳ more than one are possible, but they share the same $V_{\pi_*}(s)$, called the optimal state-value function $V_*(s)$

$$V_*(s) \triangleq \max_{\pi} V_\pi(s) \;,\; \forall s \in \mathcal{S}$$

optimal policies also share the same optimal action-value function $\triangleq q_*$

$\pi_1 \geq \pi_2$

$V_{\pi_1}(s) \geq V_{\pi_2}(s)$

$\forall s \in \mathcal{S}$

For optimal policies π_1^* & π_2^*,

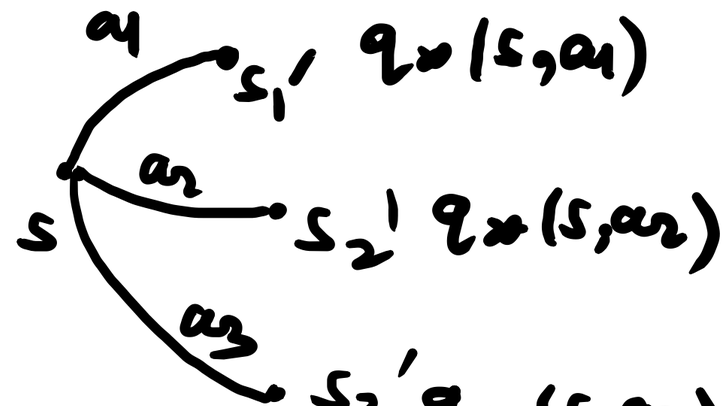
$V_{\pi_1^*}(s) = V_{\pi_2^*}(s)$

$$q_*(s,a) \doteq \max_\pi q_\pi(s,a) \qquad \forall s \in S \ \& \ \forall a \in A(s)$$

Q. $V_*(s) \stackrel{?}{=} \max_{a \in A(s)} q_{\pi_*}(s,a)$ or $q_*(s,a)$

$$V_\pi(s) = \sum_{a \in A(s)} \pi(a|s)\, q_\pi(s,a) \leq \max_{a \in A(s)} q_\pi(s,a)$$

ex.



ex2 – $c_1 x_1 + c_2 x_2 + c_3 x_3 \leq \max\{x_1, x_2, x_3\}$

where $c_1, c_2, c_3 \geq 0$ & $\sum_{i=1}^{3} c_i = 1$

$V_*(s)$

$0.2(2) + 0.3(1) + 0.5(3) <$

$0.2(3) + 0.3(3) + 0.5(3) = 3$

$$q_\pi(s,a) = E\left[G_t \mid S_t = s, A_t = a\right]$$

$$= E\left[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a\right] \quad -(5)$$

$$\overset{?}{=} E\left[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s, A_t = a\right]$$

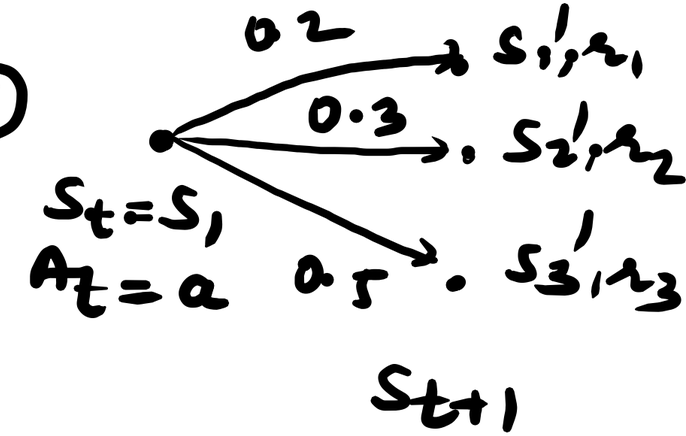
$$\gamma E_{\pi}\left[G_{t+1} \mid S_t = s, A_t = a\right]$$

$$\overset{(2)}{=} \gamma \sum_{s_{t+1}} V_\pi(s_{t+1})\, p(s_{t+1} \mid S_t = s, A_t = a)$$

$$\overset{(1)}{=} \gamma \sum_{s_{t+1}} E_\pi\left[G_{t+1} \mid S_t = s, A_t = a, S_{t+1}\right] p\left[S_{t+1} \mid S_t = s, A_t = a\right]$$

$$\overset{(3)}{=} \gamma \sum_{s_{t+1}} \sum_{r} V_\pi(s_{t+1})\, p(s_{t+1}, r \mid S_t = s, A_t = a)$$

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



$p(s', r \mid s, a)$

$$E_{\pi^*}[R_{t+1} \mid S_t = s, A_t = a] = \sum_r r\, p(r \mid s, a)$$

$$= \sum_r \sum_{s'} r\, p(s', r \mid s, a)$$

Going back to eq. (5)

$$E_{\pi^*}\left[R_{t+1} + \gamma\, q_{t+1}(S_{t+1}) \mid S_t = s, A_t = a\right]$$

$$= \sum_{s_{t+1}} \sum_r \left(r + \gamma\, q_{t+1}(s_{t+1})\right) p\left(s_{t+1}, r \mid s, a\right)$$

$$E[f(x) \mid Y]$$

$$\int_x f(x)\, p(x \mid y)$$