110652019 林楷傑

1, (a)

① $V_*(s) = \max_\pi V^\pi(s) = \max_\pi \sum_{a\in A} \pi(a|s)\, Q^\pi(s,a) \le \sum_{a\in A} \pi(a|s)\, Q_*(s,a)$

$\Rightarrow V_*(s) \le \max_a Q_*(s,a)$ ($\because$ deterministic optimal $\pi_*(a|s) = \begin{cases} 1, & a = \arg\max_{a\in A} Q_*(s,a) \\ 0, & \text{other} \end{cases}$)

Suppose $V_*(s) < \max_a Q_*(s,a)$, then $\exists \pi'$ s.t. $\exists V'_\pi(s) = \sum_{a\in A} \pi'(a|s)\, Q_*(s,a) > V_*(s)$

$\Rightarrow V'_*(s) > V_*(s) \Rightarrow (\rightarrow\leftarrow)$ with $V_*(s) = \max_\pi V^\pi(s)$

Thus, $V_*(s) = \max_a Q_*(s,a)$ #

② $Q_*(s,a) = \max_\pi Q^\pi(s,a) = \max_\pi \left[R_s^a + \gamma \sum_{s'} P_{ss'}^a V^\pi(s') \right]$

$= R_s^a + \gamma \sum_{s'} P_{ss'}^a \max_\pi V^\pi(s') = R_s^a + \gamma \sum_{s'} P_{ss'}^a V_*(s')$ #

(b)

For any two action-value function Q, Q',

$\|T^*(Q) - T^*(Q')\|_\infty = \max_{(s,a)} |T^*(Q)(s,a) - T^*(Q')(s,a)|$

$= \max_{(s,a)} \left| \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q(s',a') - \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q'(s',a') \right|$

$\le \max_{(s,a)} \left| \gamma \sum_{s'} P_{ss'}^a \max_{a'} [Q(s',a') - Q'(s',a')] \right|$

$\le \max_{(s,a)} \max_{a'} \left| \gamma \sum_{s'} P_{ss'}^a (Q(s',a') - Q'(s',a')) \right|$

$\le \gamma \|Q - Q'\|_\infty \Rightarrow T^*$ is a γ-contraction operator

2,

Let $\mu \in \mathbb{R}$, consider Largrange fuction,

$$L(\pi) = \sum_{a\in A} \left(\pi(a|s)\, Q_{\Omega}^{\pi_k}(s,a) - \pi(a|s) \log \pi(a|s)\right) - \mu\left(\sum_{a\in A} \pi(a|s) - 1\right)$$

$$\forall\, a \in A, \quad \frac{\partial L(\pi)}{\partial \pi(a|s)} = Q_{\Omega}^{\pi_k}(s,a) - \log \pi(a|s) - 1 - \mu \Rightarrow \pi(a|s) = \exp(Q_{\Omega}^{\pi_k}(s,a) - 1 - \mu)$$

$$\sum_{a\in A} \pi(a|s) = 1 \Rightarrow \sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a) - 1 - \mu) = e^{-1-\mu} \sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a)) = 1$$

$$\Rightarrow e^{1+\mu} = \sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a)) \Rightarrow \mu = \ln \sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a)) - 1$$

$$\pi(a|s) = \exp\left(Q_{\Omega}^{\pi_k}(s,a) - 1 - \left(\ln \sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a)) - 1\right)\right)$$

$$= \frac{\exp(Q_{\Omega}^{\pi_k}(s,a))}{\sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a))} \Rightarrow \text{optimal}$$

Thus, $\pi_{k+1}(\cdot|s) = \arg\max_{\pi} \left\{ \langle \pi(\cdot|s), Q_{\Omega}^{\pi_k}(s,a) \rangle - \Omega(\pi(\cdot|s)) \right\}$

$$= \frac{\exp(Q_{\Omega}^{\pi_k}(s,\cdot))}{\sum_{a\in A} \exp(Q_{\Omega}^{\pi_k}(s,a))}$$