RL HW0 、311555031、黃伯鈞

1.(a)假設 $s_0 \in S$、以及 $a^* = \text{argmax}_{a \in A(s_0)} q_*(s_0, a)$、故針對所有 policy π:

$$v_\pi(s_0) = \sum_{a \in A(s_0)} \pi(a|s_0) * q_\pi(s_0, a)$$

$$\leq \sum_{a \in A(s_0)} \pi(a|s_0) * q_*(s_0, a) \quad \left(q_*(s,a) \geq q_\pi(s,a)\right)$$

$$\leq \sum_{a \in A(s_0)} \pi(a|s_0) * q_*(s_0, a^*) \quad \text{在 optimal action 下得 } q_*(s, a^*)$$

$$= q_*(s_0, a^*) * \sum_{a \in A} \pi(a|s_0) \quad \left(\sum \pi(a|s_0) = 1\right)$$

$$= q_*(s_0, a^*) = \max_{a \in A(s_0)} q_*(s, a)$$

故對所有 policy π、可得 $v_*(s_0) = \max_\pi v_\pi(s_0) \leq \max_{a \in A(s_0)} q_*(s_0, a)$

By Contradiction、先假設 $v_*(s_0) < \max_{a \in A(s_0)} q_*(s_0, a)$

故依據假設可推得存在一 policy ϕ_t 使得 $v_{\phi_t}(s_0) = \max_{a \in A(s_0)} q_{\phi_t}(s_0, a) > v_*(s_0)$

但與 $v_*(s_0) = \max_\pi v_\pi(s_0)$ 的前提產生矛盾!!

故 $v_*(s_0) = \max_{a \in A(s_0)} q_*(s_0, a)$, for all $s \in S$ a

在代入 $q_\pi(s,a) = R_s^a + \gamma \sum_{s'} p_{ss'}^a v_\pi(s')$ 可得 $q_*(s,a) = R_s^a + \gamma \sum_{s'} p_{ss'}^a v_*(s)$

1. (b) For any two action-value Q, Q', $\|T^*(Q)-T^*(Q')\|_\infty = \max_{(s,a)} \left|[T^*Q](s,a)-[T^*Q'](s,a)\right|$

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

可得 $\|T^*(Q)-T^*(Q')\|_\infty = R_s^a + \gamma \sum_s P_{ss'}^a \left(\max Q_*(s',a')\right) - R_s^a - \gamma \sum_s P_{ss'}^a \left(\max Q'_*(s',a')\right)$

$$= \gamma \sum_{s'} P_{ss'}^a \times \left[\max Q_*(s',a') - \max Q'_*(s',a')\right]$$

故可推得 T^* is a γ-Contraction operator in terms of ∞-norm.

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

$\frac{\partial L(\pi)}{\partial \pi(a|s)} = Q_\Omega^{\pi_k}(s,a) - (\log \pi(a|s) \cdot 1 - M = 0$ for all $a \in A$

兩邊取"exp" $\Rightarrow \exp Q_\Omega^{\pi_k}(s,a) \times \frac{1}{\pi(a|s)} \times \frac{1}{e} \times \frac{1}{\exp(M)} = 1$

得到 $\frac{\exp Q_\Omega^{\pi_k}(s,a)}{e \times \exp(M)} = \pi(a|s)$ — ①

又 $\sum \pi(a|s) = 1$ 可得 $\sum \frac{\exp Q_\Omega^{\pi_k}(s,a)}{e \times \exp(M)} = 1$ — ②

由 ①/② 可得 $\frac{\exp Q_\Omega^{\pi_k}(s,a)}{\sum \exp Q_\Omega^{\pi_k}(s,a)} = \pi(a|s)$ for all $a \in A$