note

Dongkun Zhang

May 2021

1 Markov Decision Process

$$r_{t_0}^{\gamma} = \sum_{t=t_0}^{\infty} \gamma^{t-t_0} r(s_t, a_t)$$
$$J(\pi) = \mathbb{E}[r_0^{\gamma}; \pi]$$

$$J(\pi) = \mathbb{E}_{s\sim\rho^{\pi}(\cdot),\ a\sim\pi(\cdot|s)}[r(s,a)]$$

Optimization Problem:

$$\max_{\pi} J(\pi)$$

$$V^{\pi}(s) = \mathbb{E}[r_t^{\gamma}|S_t = s; \pi]$$
$$Q^{\pi}(s,a) = \mathbb{E}[r_t^{\gamma}|S_t = s, A_t = a; \pi]$$

2 Value Function

$$V^{\pi}(s) = \mathbb{E}_{a\sim\pi(\cdot|s)}[Q^{\pi}(s,a)]$$
$$Q^{\pi}(s,a) = r(s,a) + \mathbb{E}_{s'\sim p(\cdot|s,a)}[V^{\pi}(s')]$$

$$V(s) = V^*(s) = \max_{\pi} V^{\pi}(s)$$
$$Q(s,a) = Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a)$$

$$V(s) = \max_{a} Q^{\pi}(s,a)$$

Value Function *Something*

$$
\begin{aligned}
Q(s,a) &= r(s,a) + \mathbb{E}_{s' \sim p(\cdot|s,a)}[V(s')] \\
&= r(s,a) + \mathbb{E}_{s' \sim p(\cdot|s,a)}[\max_{a'} Q(s',a')]
\end{aligned}
$$

Proof.
Before

$$
\begin{aligned}
Q(s,a) &= \max_{\pi} Q^{\pi}(s,a) \\
&= r(s,a) + \max_{\pi} \mathbb{E}_{s' \sim p(\cdot|s,a)}[V(s')]
\end{aligned}
$$

After. ■

(Definition) Advantage Function *Something*

$$
\begin{aligned}
A^{\pi}(s,a) &= Q^{\pi}(s,a) - V^{\pi}(s) \\
\mathbb{E}_{a \sim \pi(\cdot|s)}[A^{\pi}(s,a)] &= 0
\end{aligned}
$$

Optimal Advantage Function *Something*

$$
\begin{aligned}
A(s,a) &= Q(s,a) - V(s) \\
A(s,a^*) = 0, \quad a^* &= \arg\max_{a} Q(s,a)
\end{aligned}
$$

Proof.
Before

$$
\begin{aligned}
V^{\pi}(s) &= \mathbb{E}_{a \sim \pi(\cdot|s)}[Q^{\pi}(s,a)] \\
&= \mathbb{E}_{a \sim \pi(\cdot|s)}[A^{\pi}(s,a) + V^{\pi}(s)] \\
&= \mathbb{E}_{a \sim \pi(\cdot|s)}[A^{\pi}(s,a)] + V^{\pi}(s)
\end{aligned}
$$

After. ■