強化学習導入

西浦学

東京大学

July N, 2020

1. Characters in RL world
2. policy gradient theorem
3. Relationship between $Q(s, a)$ function and $V(s)$

Value function

$$
\begin{aligned}
V^{\pi}(s) &= \mathbb{E}^{\pi}\left[G_t | S_t = s\right] \\
&= \sum_{s' \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} P\left(S_{t+1} = s', A_t = a | S_t = s\right) r\left(s, a, s'\right) \\
&= \sum_{s' \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} P\left(S_{t+1} = s' | S_t = s, A_t = a\right) \pi(a|s) r\left(s, a, s'\right)
\end{aligned}
$$

an example of finite horizon return $T = 2$

$$G_t = R_{t+1} + R_{t+2} \tag{1}$$

therefore the value function can be calculated as following.

$$\begin{aligned}
V^{\pi}(s) =& \mathbb{E}^{\pi}\left[G_t | S_t = s\right] = \mathbb{E}^{\pi}\left[R_{t+1} + R_{t+2} | S_t = s\right] \\
&= \sum_{s'' \in \mathcal{S}} \sum_{a' \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} \\
&P\left(S_{t+2} = s'', A_{t+1} = a', S_{t+1} = s', A_t = a | S_t = s\right) \\
&\times \left\{r\left(s, a, s'\right) + r\left(s', a', s''\right)\right\} \\
&= \sum_{s'' \in \mathcal{S}} \sum_{a' \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} \\
&P\left(S_{t+2} = s'' | S_{t+1} = s', A_{t+1} = a'\right) \pi\left(a' | s'\right) \\
&\times P\left(S_{t+1} = s' | S_t = s, A_t = a\right) \pi(a|s) \left\{r\left(s, a, s'\right) + r\left(s', a', s''\right)\right\}
\end{aligned}$$

Theorem 2.1

定理型環境が使える。使い方は普通の LaTeX と同じ

Theorem 2.1

定理型環境が使える。使い方は普通の LaTeX と同じ

Proof.

証明も書ける。 □

Theorem 2.1

定理型環境が使える。使い方は普通の LaTeX と同じ

Proof.

証明も書ける。 □

Example 2.2

example

文字の色を変えてみよう

赤

文字の色を変えてみよう

赤青

文字の色を変えてみよう

赤青緑