Solutions to Reinforcement Learning by Sutton Chapter 11

Yifan Wang

Jan 2020

Exercise 11.1

It's generally similar with (11.6) , with same treatment of the ends of episodes:

$$\mathrm{w}_{t+n} \doteq \mathrm{w}_{t+n-1} + \alpha\rho_{t:t+n-1}[G_{t:t+n} - \hat{v}(S_t, \mathrm{w}_{t+n-1})]\nabla\hat{v}(S_t, \mathrm{w}_{t+n-1})$$

$$G_{t:t+n} \doteq \sum_{k=1}^{n} \gamma^{k-1} R_{t+k} \; + \; \gamma^n \hat{v}(S_{t+n}, \mathrm{w}_{t+n-1}) \quad \text{(episodic)}$$

$$G_{t:t+n} \doteq \sum_{k=1}^{n} [R_{t+k} - \bar{R}_{t+k-1}] \; + \; \hat{v}(S_{t+n}, \mathrm{w}_{t+n-1}) \quad \text{(continuous)}$$

■

Exercise 11.2

It's still similar with SARSA but with the addition of a linear slider of σ.

$$\mathrm{w}_{t+n} \doteq \mathrm{w}_{t+n-1} + \alpha\rho_{t:t+n-1}[G_{t:t+n} - \hat{q}(S_t, \mathrm{w}_{t+n-1})]\nabla\hat{q}(S_t, \mathrm{w}_{t+n-1})$$

$$G_{t:h} \doteq R_{t+1} + \gamma\Big(\sigma_{t+1}\rho_{t+1} + (1-\sigma_{t+1})\pi(A_{t+1}|S_{t+1})\Big)\Big(G_{t+1:h} - \hat{q}(S_{t+1}, A_{t+1}, \mathrm{w}_{h-1})\Big)$$
$$+ \gamma\bar{V}_{h-1}(S_{t+1}) \quad \text{(episodic)}$$

$$G_{t:h} \doteq R_{t+1} - \bar{R}_t + \Big(\sigma_{t+1}\rho_{t+1} + (1-\sigma_{t+1})\pi(A_{t+1}|S_{t+1})\Big)\Big(G_{t+1:h} - \hat{q}(S_{t+1}, A_{t+1}, \mathrm{w}_{h-1})\Big)$$
$$+ \bar{V}_{h-1}(S_{t+1}) \quad \text{(continuous)} \qquad \text{with } \bar{V}_t(s) \doteq \sum_a \pi(a|s)\hat{q}(s, a, w_t)$$

■

Exercise 11.3

A quick solution. It's a quick and shallow solution. For better quality, one could do much better than this (like adding graph) .

■

Exercise 11.4

Following the hint, we have:

$$
\begin{aligned}
\overline{\mathrm{RE}}(w) &= \mathbb{E}\big[\big(G_t - \hat{v}(S_t, w)\big)^2\big] \\
&= \sum_s \mu(s)\big(G_t - \hat{v}(s, w)\big)^2 \\
&= \sum_s \mu(s)\big([G_t - v^*(s)] + [v^*(s) - \hat{v}(s, w)]\big)^2 \\
&= \sum_s \mu(s)\Big([G_t - v^*(s)]^2 + [v^*(s) - \hat{v}(s, w)]^2 \\
&\qquad + 2\,[G_t - v^*(s)][v^*(s) - \hat{v}(s, w)]\Big)
\end{aligned}
$$

since in expectation true value v* will equal to G, last term is zero

$$
\begin{aligned}
&= \sum_s \mu(s)\Big([G_t - v^*(s)]^2 + [v^*(s) - \hat{v}(s, w)]^2\Big) \\
&= \overline{\mathrm{VE}}(\mathrm{w}) + \mathbb{E}\Big[\big(G_t - v_\pi(S_t)\big)^2\Big].
\end{aligned}
$$

■