Rubén Moreno-Bote

© Neuro 2024

Beijing

Bellman eq

$$V_\pi(s) = \mathbb{E}_{\substack{a\sim\pi \\ s\sim p}}\left[\sum_{t=0}^{\infty}\gamma^t \underbrace{r(s_t,a_t)}_{r_{t+1}}\right] \qquad z = (s=s_0, a_0, \underbrace{s_1, a_1}_{z_1|s_1}, s_2, \dots) \quad z \text{ ends at } s$$

$$p(z|s_0) = \pi(a_0|s_0)\, p(s_1|s_0,a_0)\,\pi(a_1|s_1)\, p(s_2|s_2,a_2)\dots$$

$$V_\pi(s_0) = \sum_z p(z|s_0)\sum_{t=0}^{\infty}\gamma^t r(s_t,a_t)$$

$$= \sum_z p(z|s_0)\left[r(s_0,a_0) + \gamma r(s_1,a_1) + \gamma^2 r(s_2,a_2) + \cdots\right]$$

$$= \underbrace{\sum_z p(z|s_0) r(s_0,a_0)}_{\sum_{a_0}\pi(a_0|s_0)\, r(s_0,a_0)} + \gamma \underbrace{\sum_z p(z|s_0)\left[r(s_1,a_1) + \gamma r(s_2,a_2) + \cdots\right]}$$

$$\sum_{a_0,s_1,z_1}\pi(a_0|s_0)\, p(s_1|s_0,a_0)\, p(z_1|s_1)\left[r(s_1,a_1) + \gamma r(s_2,a_2) + \cdots\right]$$

$$\sum_{a_0,s_1}\pi(a_0|s_0)\, p(s_1|s_0,a_0)\underbrace{\sum_{z_1} p(z_1|s_1)\left[r(s_1,a_1) + \gamma r(s_2,a_2) + \cdots\right]}_{V_\pi(s_1)}$$

$s_0 \to s$, $s_1 \to s'$

$$\Rightarrow \boxed{V_\pi(s) = \sum_a \pi(a|s)\left[r(s,a) + \gamma\sum_{s'} p(s'|s,a)\, V_\pi(s')\right] = \mathop{\mathbb{E}}_{a|s\sim\pi}\left[r(s,a)\right] + \gamma \mathop{\mathbb{E}}_{\substack{a|s\sim\pi \\ s'|s,a\sim p}}\left[V_\pi(s')\right]}$$

Bellman equation

optimality Bellman equation

$$V^*(s) = \max_a\left[r(s,a) + \gamma\sum_{s'} p(s'|s,a)\, V^*(s')\right]$$

if we know $V^*(s')$ $\forall s'$ children of s

$\pi(a|s)$: $s \to a \to s'$, $V^*(s')$. Assume $s'(s,a)$, deterministic

$$V^*(s) = \max_a\left[r(s,a) + \gamma V^*(s'(s,a))\right]$$

→ major emphasis that first known the optimality Bellman eq.

How do we find $V^*(s)$?

$0 \le r(s,a) < \infty$, start with trivial guess $V_0(s) = 0$ $\forall s$

$$V_1(s) = \max_a\left[r(s,a) + \gamma\sum_{s'} p(s'|s,a)\, V_0(s')\right] = \max_a\left[r(s,a)\right] \ge 0$$

$\Rightarrow V_1(s) \ge V_0(s)$ $\forall s$

Assume $V_t(s) \ge V_{t-1}(s)$.

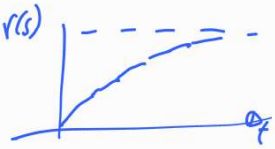
To prove that $V_{t+1}(s) \ge V_t(s)$?

$$V_{t+1}(s) = \max_a\left[r(s,a) + \gamma\sum_{s'} p(s'|s,a)\, V_t(s')\right]$$

$$\ge \max_a\left[r(s,a) + \gamma\sum_{s'} p(s'|s,a)\, V_{t-1}(s')\right] = V_t(s)$$

By induction $\Rightarrow V_{t+1}(s) \ge V_t(s)$ $\forall s, t$

↓ non-decreasing.



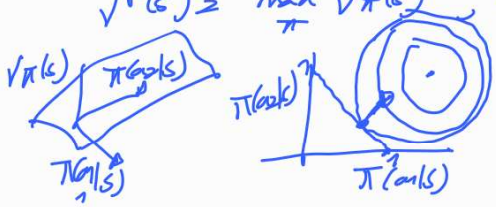
$V^*(s)$ is bounded for $\gamma < 1$

$$\text{i.e. } V^*(s) \le \frac{r_{\max}}{1-\gamma}$$

Optimal value function and Policy in Soft RL

$$V_\pi(s) = \sum_a \pi(a|s)\left[r(s,a) - \alpha \log \frac{\pi(a|s)}{\pi_0(a|s)} + \gamma \sum_{s'} p(s'|s,a) V_\pi(s')\right]$$

$$V^*(s) \geq \max_\pi V_\pi(s) \qquad \pi^* = \arg\max_\pi V_\pi(s)$$



$$\frac{\partial V_\pi(s)}{\partial \pi(a_1|s)} = \ell(s) \quad \forall a$$

$$\frac{\partial V_\pi(s)}{\partial \pi(a_2|s)} = \ell(s)$$

$$\vdots$$

Critical π

$$\frac{\partial V_\pi(s)}{\partial \pi(a|s)} = r(s,a) - \alpha \log \frac{\pi(a|s)}{\pi_0(a|s)} - \alpha \frac{\pi(a|s)}{\pi(a|s)} + \gamma \sum_{s'} p(s'|s,a) V_\pi(s') + \gamma \sum_{a'} \pi(a'|s) \sum_{s'} p(s'|s,a) \underbrace{\frac{\partial V_\pi(s')}{\partial \pi(a|s)}}_{\ell(s'|s)} = \ell(s)$$

$\alpha \frac{\pi(a|s)}{\pi(a|s)} = 1$

$\gamma \sum_{a'} \pi(a'|s) \sum_{s'} p(s'|s,a) \frac{\partial V_\pi(s')}{\partial \pi(a|s)}$ independent of a

$$\Rightarrow \boxed{\pi^*(a|s) \propto \pi_0(a|s)\, e^{\frac{1}{\alpha}\left[r(s,a) + \gamma \sum_{s'} p(s'|s,a) V^\pi(s')\right]}}$$

$Q(s,a)$

critical $\Rightarrow$ $\Leftarrow$ (critical points of boundaries of simplex)

uniqueness