PG proof

Thursday, October 17, 2019 2:11 PM

$$\nabla v^{\pi} = \nabla_{\theta} v^{\pi_\theta}$$

$$\nabla V^{\pi}(s) = \nabla \left(\sum_a \pi(a|s) Q^{\pi}(s,a) \right)$$

$$= \sum_a \nabla \left(\pi(a|s) Q^{\pi}(s,a) \right)$$

$$= \sum_a \left[\left(\nabla \pi(a|s) \right) Q^{\pi}(s,a) + \pi(a|s) \nabla Q^{\pi}(s,a) \right]$$

$$= \sum_a \pi(a|s) \frac{\nabla \pi(a|s)}{\pi(a|s)} Q^{\pi}(s,a) + \sum_a \pi(a|s) \nabla Q^{\pi}(s,a)$$

$$= \sum_a \pi(a|s) \left(\nabla \log \pi(a|s) \right) Q^{\pi}(s,a) + \sum_a \pi(a|s) \nabla \left(R(s,a) + \gamma \mathbb{E}_{s' \sim P(s,a)} \left[V^{\pi}(s') \right] \right)$$

($\nabla R(s,a) \hookrightarrow 0$.)

$$= \mathbb{E}_{a \sim \pi(s)} \left[\left(\nabla \log \pi(a|s) \right) \cdot Q^{\pi}(s,a) \right] + \gamma \sum_a \pi(a|s) \mathbb{E}_{s' \sim P(s,a)} \left[\nabla V^{\pi}(s') \right].$$

$$\nabla V^{\pi}(s) = \mathbb{E}_{a \sim \pi(s)} \left[\left(\nabla \log \pi(a|s) \right) \cdot Q^{\pi}(s,a) \right] + \gamma \mathbb{E}_{a \sim \pi(s),\, s' \sim P(s,a)} \left[\nabla V^{\pi}(s'') \right].$$

d_{π}^{t}: distribution of S_t under π. $d_{\pi}^{1} = \mu$ (initial distribution).

$$\nabla v^{\pi} = \mathbb{E}_{s\sim\mu}[\nabla V^{\pi}(s)] = \mathbb{E}_{s\sim d_{\pi}^{1}}[\nabla V^{\pi}(s)].$$

$$= \mathbb{E}_{s\sim d_{\pi}^{1}}\left[\mathbb{E}_{a\sim\pi(s)}[\square] + \gamma\,\mathbb{E}_{a\sim\pi(s),\, s'\sim P(s,a)}[\nabla V^{\pi}(s')]\right]$$

$$= \mathbb{E}_{s\sim d_{\pi}^{1},\, a\sim\pi(s)}[\square] + \gamma\,\mathbb{E}_{s\sim d_{\pi}^{1},\, a\sim\pi(s),\, s'\sim P(s,a)}[\nabla V^{\pi}(s')].$$

($s' \sim d_{\pi}^{2}$)

$$= \mathbb{E}_{s\sim d_{\pi}^{1},\, a\sim\pi(s)}[\square] + \gamma\,\mathbb{E}_{s'\sim d_{\pi}^{2}}[\nabla V^{\pi}(s')].$$

$$= \mathbb{E}_{s\sim d_{\pi}^{1},\, a\sim\pi(s)}[\square] + \gamma\,\mathbb{E}_{s\sim d_{\pi}^{2},\, a\sim\pi(s)}[\square] + \gamma^{2}\,\mathbb{E}_{s\sim d_{\pi}^{3},\, a\sim\pi(s)}[\square]$$

$$+ \cdots +$$

(recall $d_{\pi} = \sum_{t=1}^{\infty} \gamma^{t-1} d_{\pi}^{t}$. $\eta_{\pi} = (1-\gamma) d_{\pi}$.)

$$= \frac{1}{1-\gamma}\,\mathbb{E}_{s\sim\eta^{\pi},\, a\sim\pi(s)}[\nabla \log \pi(a|s)\, Q^{\pi}(s,a)].$$