Actor-Critic Methods for Control Problems

$$\nabla_\theta J(\theta) = \mathsf{E}\left[\left(\int_0^T L(X_t,u_t)\,\mathrm{d}t + \Phi(X_T)\right)\cdot\sum_{i=0}^{N-1}\log\pi_\theta(u_i \mid t_i, X_{t_i})\right]$$

$$C_t := \int_t^T L(X_s,u_s)\,\mathrm{d}s + \Phi(X_T) \qquad A^{\pi_\theta} := Q^{\pi_\theta}(t,X_t,u_t) - V^{\pi_\theta}(t,X_t)$$

$$\begin{aligned}\nabla_\theta J(\theta) &= \mathsf{E}\left[\left(\mathsf{E}\left[C_t \mid X_t,u_t\right] - V(t,X_t)\right)\cdot\sum_{i=0}^{N-1}\nabla_\theta\log\pi_\theta(u_i)\right]\\ &= \mathsf{E}\left[\sum_{i=0}^{N-1}\nabla_\theta\log\pi_\theta(u_i)\cdot A^{\pi_\theta}(t_i,X_{t_i},u_i)\right]\end{aligned}$$

$$V^{\pi_\theta}(t_i,X_{t_i}) \approx V_\phi(t_i,X_{t_i}) \qquad Q_n(t_i,X_{t_i},u_i) \approx Q^{\pi_\theta}(t_i,X_{t_i},u_i)$$

$$\begin{aligned}Q_n(t_i,X_{t_i},u_i) &:= \int_{t_i}^{t_i+n\cdot\Delta t} L(X_t,u_t)\,\mathrm{d}t + V_\phi(t_i+n\cdot\Delta t, X_{t_{i+n\cdot\Delta t}})\\ &\approx \sum_{k=0}^{n-1} C_{k+i} + V_\phi(t_{i+n},X_{t_{i+n}})\end{aligned}$$

$$\mathsf{E}\left[\underbrace{\sum_{k=0}^{N-1} C_{i+k} + V_\phi(t_{i+N},X_{t_{i+N}})}_{Q_n(t_i,X_{t_i},u_i)} \;\middle|\; X_{t_i}=x\right] \approx V_\phi(t_i,X_{t_i})$$

Erfüllt V_ϕ die Bellmanngleichung so gilt:

$$V_\phi^*(t_i,X_{t_i}) = \mathsf{E}\left[Q_n \mid X_{t_i}=x\right] \Leftrightarrow \min_\phi \mathsf{E}\left[(V_\phi - Q_n)^2\right]$$

Solving a (Stochastic) Control Problem

$$\min_\pi \mathsf{E}\left[\int_0^1 X_t + u_t^2\,\mathrm{d}t + X_1^2\right] \qquad \mathrm{d}X_t = (X_t+u_t+1)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad u_t\sim\pi_\theta(\cdot\mid t)$$

Further Directions

Further improvements based on [3], [2] and [1].

References

[1] Tuomas Haarnoja et al. "Soft Actor-Critic Algorithms and Applications". In: *arXiv preprint arXiv:1812.05905* (2018). arXiv: `1812.05905 [cs.LG]`.

[2] Tuomas Haarnoja et al. "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor". In: *Proceedings of the 35th International Conference on Machine Learning (ICML)*. 2018. arXiv: `1801.01290 [cs.LG]`.

[3] Timothy P. Lillicrap et al. "Continuous control with deep reinforcement learning". In: *arXiv preprint arXiv:1509.02971* (2015). arXiv: `1509.02971 [cs.LG]`.