15-859 Algorithms for Big Data Assignment 1

Fan Pu Zeng
fzeng@andrew.cmu.edu

1: Scratcy Scratch

$$\begin{aligned}\nabla_\theta \mathop{\mathbf{E}}_{\tau\sim P_\theta(\tau)}[R(\tau)] &= \nabla_\theta \sum_\tau P_\theta(\tau)R(\tau) \\ &= \sum_\tau \nabla_\theta P_\theta(\tau)R(\tau) \qquad \text{(uh oh..)}\end{aligned}$$

$$\begin{aligned}\sum_\tau \nabla_\theta P_\theta(\tau)R(\tau) &= \sum_\tau \frac{P_\theta(\tau)}{P_\theta(\tau)}\nabla_\theta P_\theta(\tau)R(\tau) \\ &= \sum_\tau P_\theta(\tau)\frac{\nabla_\theta P_\theta(\tau)}{P_\theta(\tau)}R(\tau) \\ &= \sum_\tau P_\theta(\tau)\nabla_\theta \log P_\theta(\tau)R(\tau) \\ &= \mathop{\mathbf{E}}_{\tau\sim P_\theta(\tau)}[\nabla_\theta \log P_\theta(\tau)R(\tau)] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta \log P_\theta(\tau_i)R(\tau_i)\end{aligned}$$

$$\begin{aligned}\nabla_\theta \log P_\theta(\tau) &= \nabla_\theta \log \left[\prod_{t=0}^{H} P(s_{t+1} \mid s_t, a_t)\cdot \pi_\theta(a_t \mid s_t),\right] \\ &= \nabla_\theta \left[\sum_{t=0}^{H} \log P(s_{t+1} \mid s_t, a_t) + \log \pi_\theta(a_t \mid s_t)\right] \\ &= \nabla_\theta \sum_{t=0}^{H} \log \pi_\theta(a_t \mid s_t) \qquad \text{(first term does not depend on } \theta\text{, becomes zero)} \\ &= \sum_{t=0}^{H} \nabla_\theta \log \pi_\theta(a_t \mid s_t)\end{aligned}$$