MS&E 346 Assignment 16

Junting Duan

March 7, 2022

1 Problem 3

Proof. (1) The score function $\nabla_\theta \log \pi(s,a;\theta)$ can be calculated as

$$\begin{aligned}
\nabla_\theta \log \pi(s,a;\theta) &= \nabla_\theta \left(\log e^{\phi(s,a)^\top \theta} - \log \left(\sum_{b\in\mathcal{A}} e^{\phi(s,b)^\top \theta} \right) \right) \\
&= \phi(s,a) - \frac{1}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^\top\theta}} \cdot \nabla_\theta \left(\sum_{b\in\mathcal{A}} e^{\phi(s,b)^\top\theta} \right) \\
&= \phi(s,a) - \frac{1}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^\top\theta}} \cdot \sum_{b\in\mathcal{A}} e^{\phi(s,b)^\top\theta} \cdot \phi(s,b) \\
&= \phi(s,a) - \sum_{b\in\mathcal{A}} \pi(s,b;\theta) \cdot \phi(s,b).
\end{aligned}$$

(2) The action-value function approximation can be constructed as

$$\begin{aligned}
Q(s,a;w) &= \nabla_\theta \log \pi(s,a;\theta)^\top \cdot w \\
&= \phi(s,a)^\top \cdot w - \sum_{b\in\mathcal{A}} \pi(s,b;\theta) \cdot \phi(s,b)^\top \cdot w.
\end{aligned}$$

(3) It holds that

$$\begin{aligned}
\mathbb{E}_\pi[Q(s,a;w)] &= \sum_{a\in\mathcal{A}} \pi(s,a;\theta) \cdot Q(s,a;w) \\
&= \sum_{a\in\mathcal{A}} \pi(s,a;\theta) \left[\phi(s,a)^\top \cdot w - \sum_{b\in\mathcal{A}} \pi(s,b;\theta) \cdot \phi(s,b)^\top \cdot w \right] \\
&= \sum_{a\in\mathcal{A}} \pi(s,a;\theta)\phi(s,a)^\top \cdot w - \sum_{a\in\mathcal{A}} \pi(s,a;\theta) \sum_{b\in\mathcal{A}} \pi(s,b;\theta)\phi(s,b)^\top \cdot w \\
&= 0
\end{aligned}$$

□