Assignment 16

Monte Fischer

March 16, 2022

Problem 3

With $\phi(s,a), \theta$ as in the problem definition and

$$pi(s,a;\theta) = \frac{\exp \phi(s,a) \cdot \theta}{\sum_{b \in \mathcal{A}} \exp \phi(s,b) \cdot \theta}$$

we compute the score function as follows:

$$\begin{aligned}
\nabla_\theta \log \pi(s,a;\theta) &= \nabla_\theta (\phi(s,a) \cdot \theta - \log \sum_{b \in \mathcal{A}} \exp \phi(s,b) \cdot \theta) \\
&= \phi(s,a) - \nabla_\theta \log \sum_{b \in \mathcal{A}} \exp \phi(s,b) \cdot \theta \\
&= \phi(s,a) - \sum_{b \in \mathcal{A}} \frac{\phi(s,b) \exp(\phi(s,b) \cdot \theta)}{\sum_{b \in \mathcal{A}} \exp(\phi(s,b) \cdot \theta)} \\
&= \phi(s,a) - \sum_{b \in \mathcal{A}} \pi(s,b;\theta) \phi(s,b) \\
&= \phi(s,a) - \mathbb{E}_\pi[\phi(s,\cdot)]
\end{aligned}$$

We wish to choose $Q(s,a;w)$ to satisfy the constraint that $\nabla_w Q(s,a;w) = \nabla_\theta \log \pi(s,a;\theta) = \phi(s,a) - \mathbb{E}_\pi[\phi(s,\cdot)]$. If we choose $Q(s,a;w) = \phi(s,a) \cdot w - \mathbb{E}_\pi[\phi(s,\cdot) \cdot w]$ then

$$\begin{aligned}
Q(s,a;w) &= (\phi(s,a) - \mathbb{E}_\pi[\phi(s,\cdot)]) \cdot w \\
&= \sum_{i=1}^{m} \frac{\partial \log \pi(s,a;\theta)}{\partial \theta_i} w_i \\
&= \nabla_\theta \log \pi(s,a,\theta) \cdot w
\end{aligned}$$

as desired. Then immediately we have

$$\begin{aligned}
\mathbb{E}_\pi[Q(s,a;w)] &= \mathbb{E}[\phi(s,a) \cdot w - \mathbb{E}[\phi(s,\cdot) \cdot w]] \\
&= \mathbb{E}[\phi(s,a) \cdot w] - \mathbb{E}[\phi(s,a) \cdot w] \\
&= 0
\end{aligned}$$

as desired.