Problem 3 of Assignment 16

1. We first see the following:

$$\begin{aligned}
\log \pi(s,a;\theta) &= \log \frac{e^{\phi(s,a)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \\
&= \log e^{\phi(s,a)^T\theta} - \log (\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}) \\
&= \phi(s,a)^T\theta - \log (\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta})
\end{aligned} \tag{1}$$

Therefore, to get the score function, we observe the following:

$$\begin{aligned}
\nabla_\theta \log \pi(s,a;\theta) &= \phi(s,a) - \frac{\sum_{b\in\mathcal{A}}(e^{\phi(s,b)^T\theta})\phi(s,b)}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \\
&= \phi(s,a) - \sum_{b\in\mathcal{A}} \frac{e^{\phi(s,b)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,b)
\end{aligned} \tag{2}$$

The Action-Value function approximation $Q(s,a;\omega)$ such that the key constraint of the Compatible Function Approximation Theorem is satisfied:

$$Q(s,a;\omega) = \phi(s,a)^T\omega - \sum_{b\in\mathcal{A}} \frac{e^{\phi(s,b)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,b)^T\omega \tag{3}$$

To verify that $Q(s,a;\omega)$ has zero mean, we see the following:

$$\begin{aligned}
\sum_{a\in\mathcal{A}} \pi(s,a;\theta)\cdot Q(s,a;\omega) &= \sum_{a\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot (\phi(s,a)^T\omega - \sum_{b\in\mathcal{A}} \frac{e^{\phi(s,b)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,b)^T\omega) \\
&= \sum_{a\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,a)^T\omega - \sum_{b\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta} e^{\phi(s,b)^T\theta}}{(\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta})^2} \cdot \phi(s,b)^T\omega \\
&= \sum_{a\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,a)^T\omega - \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta} e^{\phi(s,b)^T\theta}}{(\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta})^2} \cdot \phi(s,b)^T\omega \\
&= \sum_{a\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,a)^T\omega - \sum_{b\in\mathcal{A}}\sum_{a\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta} e^{\phi(s,b)^T\theta}}{(\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta})^2} \cdot \phi(s,b)^T\omega \\
&= \sum_{a\in\mathcal{A}} \frac{e^{\phi(s,a)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,a)^T\omega - \sum_{b\in\mathcal{A}} \frac{e^{\phi(s,b)^T\theta}}{\sum_{b\in\mathcal{A}} e^{\phi(s,b)^T\theta}} \cdot \phi(s,b)^T\omega \\
&= 0
\end{aligned} \tag{4}$$