Jordan Rowley

CME 241: Assignment 16

16.3

(a) Since

$$\log(\pi(s,a;\boldsymbol{\theta})) = \phi(s,a)^T\boldsymbol{\theta} - \log\left(\sum_{b\in\mathcal{A}} \exp\left(\phi(s,b)^T\boldsymbol{\theta}\right)\right)$$

we have that

$$\nabla_{\boldsymbol{\theta}} \log(\pi(s,a;\boldsymbol{\theta})) = \phi(s,a) - \frac{\sum_{b\in\mathcal{A}} \exp\left(\phi(s,b)^T\boldsymbol{\theta}\right)\cdot\phi(s,b)}{\sum_{b\in\mathcal{A}} \exp\left(\phi(s,b)^T\boldsymbol{\theta}\right)} = \phi(s,a) - \sum_{b\in\mathcal{A}} \pi(s,b;\boldsymbol{\theta})\phi(s,b)$$

(b) To construct a State-Action function $Q(s,a;\boldsymbol{w})$ for which $\nabla_{\boldsymbol{w}} Q = \nabla_{\boldsymbol{\theta}} \log(\pi(s,a;\boldsymbol{\theta}))$, we can simply set

$$Q(s,a;\boldsymbol{w}) = \nabla_{\boldsymbol{\theta}} \log(\pi(s,a;\boldsymbol{\theta}))^T\boldsymbol{w}$$

so that

$$Q(s,a;\boldsymbol{w}) = u^T\boldsymbol{w}$$

where $u = \phi(s,a) - \sum_{b\in\mathcal{A}} \pi(s,b;\boldsymbol{\theta})\phi(s,b)$

(c) For a given state s, the mean of Q is equal to

$$\mathbb{E}_\pi[Q(s,a;\boldsymbol{w})] = \sum_{a\in\mathcal{A}} \pi(s,a;\boldsymbol{\theta})\cdot Q(s,a;\boldsymbol{w})$$

$$= \sum_{a\in\mathcal{A}} \pi(s,a,\boldsymbol{\theta})\left(\phi(s,a) - \sum_{b\in\mathcal{A}} \pi(s,b;\boldsymbol{\theta})\phi(s,b)\right)$$

$$= \sum_{a\in\mathcal{A}} \pi(s,a,\boldsymbol{\theta})\phi(s,a) - \sum_{b\in\mathcal{A}} \pi(s,b;\boldsymbol{\theta})\phi(s,b) = 0$$

since $\sum_{a\in\mathcal{A}} \pi(s,a,\boldsymbol{w}) = 1$. Therefore, $\mathbb{E}_\pi[Q(s,a;\boldsymbol{w})] = 0$