Problem 3

- Finite Action Space A
- $\phi(s,a) = \left(\phi_1(s,a), \quad \phi_2(s,a), \quad \ldots, \quad \phi_m(s,a)\right)$ feature vectors for $s \in N$, $a \in A$
- $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ parameter vector

$$\pi(s, a; \theta) = \frac{\exp\left[\phi(s,a)^T \theta\right]}{\sum_{b \in A} \exp\left[\phi(s,b)^T \theta\right]}$$

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

$$= \phi(s,a)^T\theta - \log\left(\sum_{b\in A} \exp\left[\phi(s,b)^T\theta\right]\right)$$

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

$$= \phi(s,a)^T - \frac{\nabla_\theta \sum_{b\in A} \exp\left[\phi(s,b)^T\theta\right]}{\left(\sum_{b\in A}\exp\left[\phi(s,b)^T\theta\right]\right)}$$

$$= \phi(s,a)^T - \frac{\sum_{b\in A} \nabla_\theta \exp\left[\phi(s,b)^T\theta\right]}{\left(\sum_{b\in A}\exp\left[\phi(s,b)^T\theta\right]\right)}$$

$$= \phi(s,a)^T - \frac{\sum_{b \in A} \phi(s,b)^T \exp\left[\phi(s,b)^T \theta\right]}{\left(\sum_{b \in A} \exp\left[\phi(s,b)^T \theta\right]\right)}$$

$$= \phi(s,a)^T - \underbrace{\sum_{b \in A} \phi(s,b)^T \pi(s,b;\theta)}_{\text{Weighted average over } A}$$

$$= \phi(s,a) - \mathbb{E}_A\left[\phi(s,\cdot)^T\right]$$

- **<u>Next Objective:</u>** Find $Q(s,a;w)$ such that the following condition is satisfied:

$$\nabla_w Q(s,a;w) = \nabla_\theta \log\left(\pi(s,a;\theta)\right)$$

$$\nabla_\theta \log\left(\pi(s,a;\theta)\right) = \phi(s,a)^T - \sum_{b \in A} \phi(s,b)^T \underbrace{\pi(s,b;\theta)}_{\text{scalar!}}$$

$$= \nabla_w \phi(s,a)^T w - \sum_{b \in A} \pi(s,b;\theta) \nabla_w \phi(s,b)^T w$$

$$= \nabla_w \phi(s,a)^T w - \nabla_w \sum_{b \in A} \pi(s,b;\theta) \phi(s,b)^T w$$

$$= \nabla_w \left\{ \phi(s,a)^T w - \sum_{b \in A} \pi(s,b;\theta) \phi(s,b)^T w \right\}$$

$$\therefore Q(s,a;w) = \phi(s,a)^T w - \sum_{b \in A} \pi(s,b;\theta) \phi(s,b)^T w$$

• <u>Next objective:</u> Show $\sum_{a \in A} \pi(s, a; \theta) \cdot Q(s, a; w) = 0$

$$\sum_{a \in A} \pi(s, a; \theta) \cdot Q(s, a; w) = \sum_{a \in A} \pi(s, a; \theta) \cdot \left(\phi(s, a)^T w - \sum_{b \in A} \pi(s, b; \theta) \phi(s, b)^T w \right)$$

$$= \sum_{a \in A} \pi(s, a; \theta) \phi(s, a)^T w - \sum_{a \in A} \pi(s, a; \theta) \underbrace{\sum_{b \in A} \pi(s, b; \theta) \phi(s, b)^T w}_{\substack{\text{Independent of } a \\ \therefore \text{ We can factor this out} \\ \text{of the summation}}}$$

$$= \sum_{a \in A} \pi(s, a; \theta) \phi(s, a)^T w - \sum_{b \in A} \pi(s, b; \theta) \phi(s, b)^T w \underbrace{\cancelto{1}{\sum_{a \in A} \pi(s, a; \theta)}}_{\substack{\text{Probability over all} \\ \text{actions add to } 1}}$$

$$= \sum_{a \in A} \pi(s, a; \theta) \phi(s, a)^T w - \sum_{b \in A} \pi(s, b; \theta) \phi(s, b)^T w$$

↪ These two terms are equivalent

$= 0$

∎