CME 241

Lilian Kourti

Problem 3 - Assignment 16

Let us use the softmax function to approximate the policy function

$$\pi(s,a;\theta) = \frac{e^{\varphi(s,a)^T\theta}}{\sum_{b\in A} e^{\varphi(s,b)^T\theta}}$$

Then $\log \pi(s,a;\theta) = \varphi(s,a)^T\theta - \log \sum_{b\in A} e^{\varphi(s,b)^T\theta}$

And the partial derivatives w.r.t. θ_i are given by:

$$\frac{\partial \log \pi(s,a;\theta)}{\partial \theta_i} = \varphi_i(s,a) - \frac{\sum_{b\in A}\varphi_i(s,b)e^{\varphi(s,b)^T\theta}}{\sum_{b\in A} e^{\varphi(s,b)^T\theta}}$$

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

$$= \varphi_i(s,a) - \sum_{b\in A}\pi(s,b;\theta)\varphi_i(s,b)$$

$$= \varphi_i(s,a) - E_\pi[\varphi_i(s,\cdot)]$$

Therefore

$$\nabla_\theta \log \pi(s,a;\theta) = \varphi(s,a) - E_\pi[\varphi(s,\cdot)]$$

A simple way to enable the Compatible Function Approximation $\nabla_w Q(s,a;w) = \nabla_\theta \log \pi(s,a;\theta)$ is to set $Q(s,a;w)$ to be linear in its features. We let the features of $Q(s,a;w)$ be $\nabla_\theta \log \pi(s,a;\theta)$, so we get:

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

Thus, it is easily observed that the required condition holds, since:

$$\nabla_w (w^T \nabla_\theta \log \pi(s,a;\theta)) = \nabla_\theta \log \pi(s,a;\theta)$$

$$E_\pi[Q(s,a;w)] = \sum_{a\in A}\pi(s,a;\theta)Q(s,a;w)$$

$$= \sum_{a\in A}\pi(s,a;\theta)w^T\nabla_\theta \log\pi(s,a;\theta)$$

$$= \sum_{a\in A}\pi(s,a;\theta)\sum_{i=1}^{m} w_i \frac{\partial \log \pi(s,a;\theta)}{\partial\theta_i}$$

$$= \sum_{a\in A}\pi(s,a;\theta)\sum_{i=1}^{m} w_i \frac{1}{\pi(s,a;\theta)}\frac{\partial \pi(s,a;\theta)}{\partial\theta_i}$$

$$= \sum_{a\in A}\sum_{i=1}^{m} w_i \frac{\partial \pi(s,a;\theta)}{\partial\theta_i}$$

$$= \sum_{i=1}^{m} w_i \frac{\partial}{\partial\theta_i}\left(\sum_{a\in A}\pi(s,a;\theta)\right) = \sum_{i=1}^{m} w_i \cdot \frac{\partial}{\partial\theta_i}(1) = 0$$