(1) REINFORCE on the Cart-Pole

(a) $\pi(a|s,\theta) = \dfrac{e^{h(s,a,\theta)}}{\sum_b e^{h(s,b,\theta)}}$

In the given situation,

$\pi(a|s,\theta) = \dfrac{e^{\theta_a^T s}}{e^{\theta_a^T s} + e^{\theta_b^T s}}$, where a and b two actions

Now, $s \in \mathbb{R}^4$

$$\theta = \begin{bmatrix} \theta_{a_1} & \theta_{a_2} & \theta_{a_3} & \theta_{a_4} \\ \theta_{b_1} & \theta_{b_2} & \theta_{b_3} & \theta_{b_4} \end{bmatrix}$$

$$\nabla_\theta \pi(a|s,\theta) = \frac{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)\frac{\partial}{\partial\theta}\left(e^{\theta_a^T s}\right) - e^{\theta_a^T s}\frac{\partial}{\partial\theta}\left(e^{\theta_a^T}\right.}{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)^2}$$

$$= \frac{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)e^{\theta_a^T s}\begin{bmatrix} s \\ 0 \end{bmatrix}^T - e^{\theta_a^T s}\begin{bmatrix} e^{\theta} \\ e^{\theta} \end{bmatrix}}{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)^2}$$

$$= \frac{e^{\theta_a^T s}\begin{bmatrix} \left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)s^T - e^{\theta_a^T s}\, s^T \\ 0 \quad - e^{\theta_b^T s}\, s^T \end{bmatrix}}{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)^2}$$

$$\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)^{-}$$

$$= \frac{e^{\theta_a^T s}\begin{bmatrix} e^{\theta_b^T s} s^T \\ -e^{\theta_b^T s} s^T \end{bmatrix}}{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)^2}$$

$$= \frac{e^{\theta_a^T s} e^{\theta_b^T s}\begin{bmatrix} s^T \\ -s^T \end{bmatrix}}{\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)^2}$$

$$= h(s,a,\theta) \cdot h(s,b,\theta)\begin{bmatrix} s^T \\ -s^T \end{bmatrix}$$

$$= h(s,a,\theta) \cdot h(s,b,\theta)\begin{bmatrix} s \\ -s \end{bmatrix}^T$$

(b) $\log \pi(a|s,\theta) = \log \frac{e^{\theta_a^T s}}{e^{\theta_a^T s} + e^{\theta_b^T s}}$

$$= \log\left(e^{\theta_a^T s}\right) - \log\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)$$

$$= \theta_a^T s - \log\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)$$

$\nabla_\theta \log \pi(a|s,\theta)$

$$= \nabla_\theta\left[\theta_a^T s - \log\left(e^{\theta_a^T s} + e^{\theta_b^T s}\right)\right]$$

$$\begin{bmatrix} s^T \end{bmatrix} \qquad \frac{1}{} \qquad \begin{bmatrix} e^{\theta_a^T s} \cdot s^T \end{bmatrix}$$

$$= \begin{bmatrix} s^T \\ 0 \end{bmatrix} - \frac{1}{(e^{\theta_a^T s} + e^{\theta_b^T s})} \begin{bmatrix} e^{\theta_a^T s} \cdot s^T \\ e^{\theta_b^T s} \cdot s^T \end{bmatrix}$$

$$= \begin{bmatrix} s^T - \frac{e^{\theta_a^T s}}{e^{\theta_a^T s} + e^{\theta_b^T s}} s^T \\ 0 - \frac{e^{\theta_b^T s}}{e^{\theta_a^T s} + e^{\theta_b^T s}} s^T \end{bmatrix}$$

$$= \begin{bmatrix} 1 - p(a|s,\theta)\, s^T \\ -p(b|s,\theta)\, s^T \end{bmatrix}$$

$$= \begin{bmatrix} p(b|s,\theta)\, s^T \\ -p(b|s,\theta)\, s^T \end{bmatrix}$$

$$= p(b|s,\theta) \begin{bmatrix} s^T \\ -s^T \end{bmatrix} = p(b|s,\theta) \begin{bmatrix} s \\ -s \end{bmatrix}^T$$