

1 Score Function

$$\begin{aligned}\nabla_{\theta} \log \pi(a=1|s, \theta) &= \nabla_{\theta} \log \sigma(s^T \theta) \\&= \frac{1}{\sigma(s^T \theta)} \nabla_{\theta} \sigma(s^T \theta) \\&= \frac{1}{\sigma(s^T \theta)} \sigma(s^T \theta) (1 - \sigma(s^T \theta)) s \\&= (1 - \sigma(s^T \theta)) s \\&= \pi(a=0|s, \theta) \cdot s \\ \nabla_{\theta} \log \pi(a=0|s, \theta) &= \nabla_{\theta} \log (1 - \sigma(s^T \theta)) \\&= \frac{1}{1 - \sigma(s^T \theta)} \nabla_{\theta} [1 - \sigma(s^T \theta)] \\&= -\frac{\sigma(s^T \theta)}{1 - \sigma(s^T \theta)} (1 - \sigma(s^T \theta)) s \\&= -\sigma(s^T \theta) s \\&= -\pi(a=1|s, \theta) \cdot s\end{aligned}$$

2 REINFORCE

See implementation.