

# Reinforcement Learning Lab 5 Policy gradient

Niccolò Turcato (niccolo.turcato@phd.unipd.it)

Department of Information Engineering - Università Degli studi di Padova

# Policy gradient for linear function approximation

Define:

$$h(s, a, \theta) = \theta^T x(s, a) \tag{1}$$

and consider Policy:

$$\pi(\boldsymbol{a}|\boldsymbol{s},\theta) = \frac{\boldsymbol{e}^{h(\boldsymbol{s},\boldsymbol{a},\theta)}}{\sum_{\boldsymbol{b}} \boldsymbol{e}^{h(\boldsymbol{s},\boldsymbol{b},\theta)}} \tag{2}$$

The gradient is defined as:

$$\nabla \hat{J}(\theta) = G_t \nabla_{\theta} ln(\pi(a|s,\theta)) \tag{3}$$



14/12/2022

$$abla_{ heta} \mathit{In}(\pi(\pmb{a}|\pmb{s}, heta)) = 
abla_{ heta} \mathit{In}\left(rac{\pmb{e}^{h(\pmb{s}, \pmb{a}, heta)}}{\sum_{\pmb{b}} \pmb{e}^{h(\pmb{s}, \pmb{b}, heta)}}
ight) = 
abla_{ heta}\left(\mathit{In}(\pmb{e}^{h(\pmb{s}, \pmb{a}, heta)}) - \mathit{In}(\Sigma_{\pmb{b}} \pmb{e}^{h(\pmb{s}, \pmb{b}, heta)})
ight)$$

14/12/2022

$$egin{aligned} 
abla_{ heta} In(\pi(a|s, heta)) &= 
abla_{ heta} In\left(rac{e^{h(s,a, heta)}}{\Sigma_b e^{h(s,b, heta)}}
ight) = 
abla_{ heta} \left(In(e^{h(s,a, heta)}) - In(\Sigma_b e^{h(s,b, heta)})
ight) \ &= 
abla_{ heta} h(s,a, heta) - 
abla_{ heta} In(\Sigma_b e^{h(s,b, heta)}) = x(s,a) - rac{
abla_{ heta} \Sigma_b e^{h(s,b, heta)}}{\Sigma_b e^{h(s,b, heta)}} \end{aligned}$$

14/12/2022 4/

$$\begin{split} \nabla_{\theta} \textit{In}(\pi(a|s,\theta)) &= \nabla_{\theta} \textit{In}\left(\frac{e^{h(s,a,\theta)}}{\Sigma_{b}e^{h(s,b,\theta)}}\right) = \nabla_{\theta}\left(\textit{In}(e^{h(s,a,\theta)}) - \textit{In}(\Sigma_{b}e^{h(s,b,\theta)})\right) \\ &= \nabla_{\theta} h(s,a,\theta) - \nabla_{\theta} \textit{In}(\Sigma_{b}e^{h(s,b,\theta)}) = x(s,a) - \frac{\nabla_{\theta}\Sigma_{b}e^{h(s,b,\theta)}}{\Sigma_{b}e^{h(s,b,\theta)}} \\ &= x(s,a) - \frac{\Sigma_{b}e^{h(s,b,\theta)}\nabla_{\theta}h(s,b,\theta)}{\Sigma_{b}e^{h(s,b,\theta)}} = x(s,a) - \frac{\Sigma_{b}e^{h(s,b,\theta)}x(s,b)}{\Sigma_{b}e^{h(s,b,\theta)}} \end{split}$$

14/12/2022 5/

$$\begin{split} \nabla_{\theta} \textit{In}(\pi(a|s,\theta)) &= \nabla_{\theta} \textit{In}\left(\frac{e^{h(s,a,\theta)}}{\Sigma_{b}e^{h(s,b,\theta)}}\right) = \nabla_{\theta}\left(\textit{In}(e^{h(s,a,\theta)}) - \textit{In}(\Sigma_{b}e^{h(s,b,\theta)})\right) \\ &= \nabla_{\theta} h(s,a,\theta) - \nabla_{\theta} \textit{In}(\Sigma_{b}e^{h(s,b,\theta)}) = x(s,a) - \frac{\nabla_{\theta}\Sigma_{b}e^{h(s,b,\theta)}}{\Sigma_{b}e^{h(s,b,\theta)}} \\ &= x(s,a) - \frac{\Sigma_{b}e^{h(s,b,\theta)}\nabla_{\theta}h(s,b,\theta)}{\Sigma_{b}e^{h(s,b,\theta)}} = x(s,a) - \frac{\Sigma_{b}e^{h(s,b,\theta)}x(s,b)}{\Sigma_{b}e^{h(s,b,\theta)}} \\ &= x(s,a) - \Sigma_{b}\left(\frac{e^{h(s,b,\theta)}}{\Sigma_{b}e^{h(s,b,\theta)}}\right)x(s,b) = x(s,a) - \Sigma_{b}\pi(b|s,\theta)x(s,b) \end{split}$$

□ > < □ > < □ > < □ > < □ >

14/12/2022 6/6