2

Wednesday, October 2, 2019 1:52 PM

1) Calculating $\nabla_w \log \pi_w(A_t|s_t)$

$\to$ Gradient w.r.t weights $\Rightarrow$

$$\begin{array}{c c} & a=4 \\ s=4 & \begin{bmatrix} -p(a_j|s_1)- \\ -p(a_j|s_2)- \\ -p(a_j|s_3)- \\ -p(a_j|s_4)- \end{bmatrix} \end{array}$$

1.1

$$\Rightarrow \nabla_w \left[\log\left(\frac{e^{w_{s_t,a_t}/\gamma}}{\sum_{a'} e^{w_{s_t}/a'}} \right) \right]$$

$$\Rightarrow \nabla_w \left[\underbrace{\log\left(e^{w_{s_t,a_t}/\gamma}\right)}_{i} - \underbrace{\log\left(\sum_{a'} e^{w_{s_t,a'}}\right)}_{ii} \right]$$

While W is a 4x4 matrix, we only have gradients through the specific row (representing state s_t of our choosing). Therefore :

For some i :

i) $\nabla_{w_{s_t,i}} \log(e^{w_{s_t,a_t}}/\gamma) = \nabla_{w_{s_t,i}} \frac{w_{s_t,a_t}}{\gamma} = \begin{cases} 1/\gamma & i=a_t \\ 0 & i \neq a_t \end{cases}$

ii) $\nabla_{w_{s_t,i}} \log\left(\sum_{a'} e^{w_{s_t,a'}}\right) = \frac{1}{\sum_{a'} e^{w_{s_t,a'}}} \times \frac{\partial \sum e^{w_{s_t,a'}}}{\partial w_{s_t,i}}$

$$\Rightarrow \frac{e^{w_{s_t,i}}}{\sum_{a'} e^{w_{s_t,a'}}}$$

$\therefore$ If $\gamma=1$

$$\nabla w_{s',a'} = \begin{cases} 1 - \frac{e^{w_{s_t,a_t}}}{\sum_{a''} e^{w_{s_t,a''}}} & s'=s_t,\ a'=a_t \\ \frac{-e^{w_{s_t,a_t}}}{\sum_{a''} e^{w_{s_t,a''}}} & s'=s_t,\ a'\neq a_t \\ 0 & \text{o.t.w} \end{cases}$$

2.2

2) $\pi_k(a|s) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{\frac{-(a-ks)^2}{2\sigma^2}}$

$$\log \pi_k(a|s) = \log\left(\frac{1}{\sigma\sqrt{2\pi}} \times e^{\frac{-(a-ks)^2}{2\sigma^2}} \right)$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(a-ks)^2}{2\sigma^2}$$

$\therefore \nabla_k \log \pi_k(a|s) = \frac{1}{2\sigma^2} \nabla_k (a-ks)^2$

$$\Rightarrow \frac{a-ks}{\sigma^2} \times s$$

2.1 Linear Policy

$N(s) = ks \Rightarrow s = \begin{bmatrix} x\text{-pos} \\ \text{velocity} \end{bmatrix} \quad k = 2\begin{bmatrix} 1 \\ k_1 \\ k_2 \end{bmatrix}$

Mean is a linear combination of features $k_1 \times \text{pos} + k_2\,\text{velocity}$

i) ∇