Exercices "auto. diff"

I - fonction sigmoïde / logistique

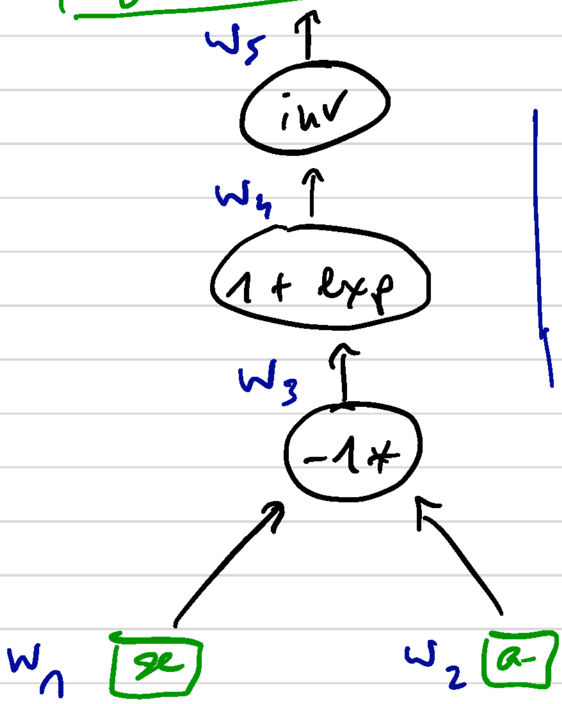
1 - tracer le graphe de calcul de cette fn :

$$y = f(x,a) = 1/(1+e^{-ax})$$

2 - trouver la dérivée en mode forward / r̄ à x : "d_x" $= \dfrac{\partial y}{\partial x}$

3 - Idem en mode reverse

$$\boxed{y = f(x,a) = 1/(1+e^{-ax})}$$



2 - Mode Forward "seeds"

$$\begin{aligned} w_1 &= x, \quad w_2 = a \\ w_3 &= -w_1 w_2 \\ w_4 &= 1 + e^{w_3} \\ w_5 &= 1/w_4 = y \end{aligned} \Rightarrow \begin{aligned} &\boxed{\dot{w}_1 = 1, \quad \dot{w}_2 = 0} \\ &\dot{w}_3 = -(\dot{w}_1 w_2 + w_1 \dot{w}_2) = -\dot{w}_1 w_2 = -w_2 \\ &\dot{w}_4 = \dot{w}_3 e^{w_3} \\ &\dot{w}_5 = -\frac{\dot{w}_4}{w_4^2} \text{ "="} d_x \end{aligned}$$

$$d_x = -\frac{\dot{w}_4}{w_4^2} = -\frac{\dot{w}_3 e^{w_3}}{w_4^2} = \frac{w_2 e^{w_3}}{w_4^2} = \dots = \frac{a e^{-ax}}{(1+e^{-ax})^2}$$

$$\boxed{d_x = \frac{a e^{-ax}}{(1+e^{-ax})^2} = \dots = a y (1-y)}$$

2 - Mode reverse : $\boxed{\text{"}d_x\text{"} = \bar{w}_1}$

$$\begin{aligned} w_1 &= x, \quad w_2 = a \\ w_3 &= -w_1 w_2 \\ w_4 &= 1 + e^{w_3} \\ w_5 &= 1/w_4 = y \end{aligned}$$

$$\bar{w}_1 = \bar{w}_3 \frac{\partial w_3}{\partial w_1} = \bar{w}_3(-w_2) = d_x \text{ "=" } \frac{a e^{-ax}}{(1+e^{-ax})^2} \text{ "=" } \dots \text{ "=" } a y (1-y)$$

$$\bar{w}_3 = \bar{w}_4 \frac{\partial w_4}{\partial w_3} = -\frac{e^{w_3}}{\bar{w}_4^2}$$

$$\bar{w}_4 = \frac{\partial y}{\partial w_4} = \bar{w}_5 \frac{\partial w_5}{\partial w_4} = -\frac{1}{w_4^2}$$

$$\bar{w}_5 = \frac{\partial y}{\partial w_5} = 1 \text{ "seed"}$$

$\bar{w}_i$: "dérivée adjointe" $= \dfrac{\partial y}{\partial w_i}$

4 - Sans faire de nouveau calcul, donner "da".

Noter

+ mode "accumulation forward" : top ↑ bottom

$$\frac{dw_i}{dx} = \frac{dw_i}{\partial w_{i-1}} \frac{dw_{i-1}}{dx} \quad \text{avec } w_n = y$$

+ mode "accu. reverse" : top ↓ bottom

$$\bar{w}_i = \frac{dy}{dw_i} = \frac{dy}{dw_{i+1}} \frac{dw_{i+1}}{dw_i}$$

↑ "adjoint"

Remarque : . en mode forward, pour calculer le gradient de $f : x_1, x_2 \mapsto f(x_1, x_2)$ on doit faire les calculs de dérivée deux fois : une fois pour x_1 et une fois pour x_2.

. en mode reverse, c'est différent : on a autant de calculs de dérivées à faire que de "sorties y" en sortie du graphe.

$\Rightarrow$ Si $f : \mathbb{R}^n \to \mathbb{R}^m$, alors

forward est plus efficace si $n \ll m$

reverse " " $n \gg m$

Remarque : l'algo backprop est un mode reverse

II Exercice 2 (maison)

On considère $\left| \begin{array}{l} f : \mathbb{R}^2 \longrightarrow \mathbb{R} \\ x_1, x_2 \longmapsto f(x_1, x_2) = x_1 x_2 + \sin(x_1) \end{array} \right.$

- Calculer "dx_1" en mode forward et reverse

Ex. MLP MSE

correction janv. 2022

1) Passe forward

$$\vec{a}_{12} = \begin{bmatrix} 0,1 & 0,1 & 0,8 \\ 0,1 & 0,4 & 0,6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1,1 \end{bmatrix}$$

$$\vec{h}_{12} = sig(\vec{a}) = \begin{bmatrix} 0,731 \\ 0,75 \end{bmatrix}$$

$$a_3 = \begin{bmatrix} 0,1 & 0,3 & 0,9 \end{bmatrix} \begin{bmatrix} 1 \\ 0,731 \\ 0,75 \end{bmatrix} = 0,9943$$

$$h_3 = \tanh(0,9943) = 0,759 = \hat{y}$$

$$loss = 0,5 * (0,5 - 0,759)^2 = 0,0335$$

2)

$$\frac{\partial loss}{\partial w_{31}} = \underbrace{\frac{\partial loss}{\partial h_3} \frac{\partial h_3}{\partial a_3}}_{=\delta_3} \frac{\partial a_3}{\partial w_{31}} = \frac{\partial}{\partial h_3}\left(0,5(y-h_3)^2\right) \times \frac{\partial \tanh(a_3)}{\partial a_3} \times \frac{\partial\, w_{31} h_1 + w_{32} h_2 + 0,1}{\partial w_{31}}$$

"dW_{31}"

$$= \underbrace{(h_3 - y)\left(1 - \tanh^2(a_3)\right)}_{\delta_3} \times h_1 = \delta_3 \times h_1$$

$$= (h_3 - y)(1 - h_3^2) \times h_1$$

$$= (0,76 - 0,5)(1 - 0,76^2) \times 0,731 = 0,080 \quad 79$$

$$\frac{\partial loss}{\partial w_{32}} = (h_3 - y)(1 - h_3^2) \times h_2 = 0,081 = \delta_3 \times h_2$$

$$\frac{\partial loss}{\partial w_{1A}} = \underbrace{\frac{\partial loss}{\partial h_1} \frac{\partial h_1}{\partial a_1}}_{=\delta_1} \frac{\partial a_1}{\partial w_{1A}} = \frac{\partial loss}{\partial a_3} \frac{\partial a_3}{\partial h_1} \times \frac{\partial h_1}{\partial a_1} \frac{\partial a_1}{\partial w_{1A}}$$

$$\cancel{= (h_3 - y)(1 - \tanh^2(}$$

$$= (h_3 - y)(1 - h_3^2)\, w_{31} \times \frac{\partial}{\partial a_1} sig(a_1) \times x_A$$

$$= \delta_3 \times w_{31} \times \left[h_1 \times (1 - h_1)\right] \times x_A = \delta_1 \times x_A$$

$$= 0,109 \times 0,3 \times 0,731 \times 0,269 \times 1 = 0,0064$$

$$\frac{\partial loss}{\partial w_{1B}} = \frac{\partial loss}{\partial h_1} \frac{\partial h_1}{\partial a_1} \frac{\partial a_1}{\partial w_{1B}} = \text{"} \times x_B = \delta_1 \times x_B = 0,0064$$

$$\frac{\partial loss}{\partial W_{2A}} = \underbrace{\delta_3 \, W_{32} \, h_2 (1 - h_2)}_{\text{"}\delta_2} x_A$$

$$= 0{,}0183$$

$$\frac{\partial loss}{\partial W_{2B}} = \delta_2 \, x_B = 0{,}0183$$

Formule générale :

neurone suivant

entrée du neurone

$$\cancel{\partial} dW_{ji} = \delta \times W_{ji} \times f' \times x$$

dérivée de l'activation

Et les biais ?

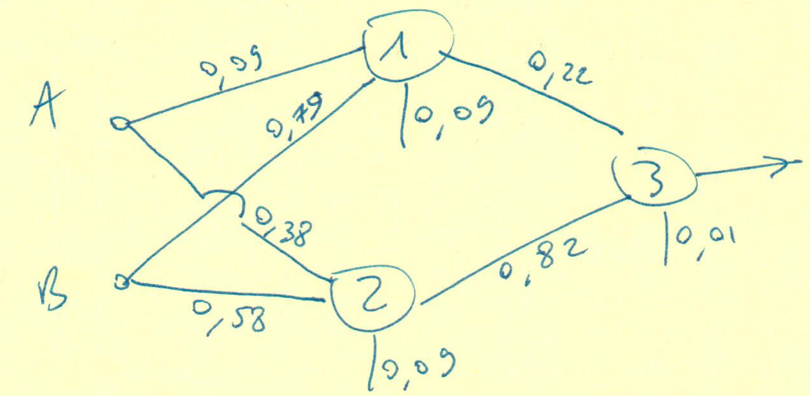
$$\frac{\partial loss}{\partial b_3} = \frac{\partial loss}{\partial h_3} \times \frac{\partial h_3}{\partial a_3} \times \frac{\partial a_3}{\partial b_3} = \text{"}db_3\text{"}$$

$$= \delta_3 \times 1 = 0{,}109$$

idem : $\frac{\partial loss}{\partial b_1} = \delta_1 = 0{,}0064$ et $\frac{\partial loss}{\partial b_2} = \delta_2 = 0{,}0183$

$$W = W - \alpha \frac{\partial loss}{\partial W}$$

3)



4) passe forward

$$\vec{a}_1 = \begin{bmatrix} 0.09 & 0.09 & 0.79 \\ 0{,}08 & 0.38 & 0.58 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.97 \\ 1.04 \end{bmatrix}$$

$$\vec{h}_1 = \begin{bmatrix} 0{,}725 \\ 0{,}738 \end{bmatrix}$$

$$a_3 = \begin{bmatrix} 0{,}01 & 0{,}22 & 0{,}82 \end{bmatrix} \begin{bmatrix} 1 \\ 0{,}725 \\ 0{,}738 \end{bmatrix} = 0{,}774$$

$h_3 = 0{,}649 \Rightarrow$ la prédiction est meilleure, plus proche de $y = 0{,}5$

$loss = 0{,}5(0{,}5 - 0{,}649)^2 = 0{,}011 < 0{,}033$

la loss est + petite.