Aprendizagem 2021/22

Homework III – Group 041

I. Pen-and-paper

1)

1-

$(0{,}8 ; 1 ; 1.2 ; 1{,}4 ; 1{,}6)$

$$\underset{\text{input}}{\begin{pmatrix} 0.8 \\ 1 \\ 1.2 \\ 1.4 \\ 1.6 \end{pmatrix}} \quad \underset{\text{output}}{\begin{pmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{pmatrix}}$$

$$\hat{z}(x, w) = \sum_{j=0}^{3} w_j \phi_j(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3$$

$\lambda = 2$ Ridge regression

Design the matrix

$$\Phi = \begin{pmatrix} 1 & 0.8 & 0{,}64 & 0{,}512 \\ 1 & 1 & 1 & 1 \\ 1 & 1{,}2 & 1{,}44 & 1{,}728 \\ 1 & 1{,}4 & 1{,}96 & 2{,}744 \\ 1 & 1{,}6 & 2{,}56 & 4{,}096 \end{pmatrix}$$

$$y(x_1|w) = 1 + 0{,}8w_1 + 0{,}64w_2 + 0{,}512w_3$$

$$y(x_2|w) = 1 + 1w_1 + 1w_2 + 1w_3$$

$$y(x_3|w) = 1 + 1{,}2w_1 + 1{,}44w_2 + 1{,}728w_3$$

$$y(x_4|w) = 1 + 1{,}4w_1 + 1{,}96w_2 + 2{,}744w_3$$

$$y(x_5|w) = 1 + 1{,}6w_1 + 2{,}56w_2 + 4{,}096w_3$$

Aprendizagem 2021/22

Homework III – Group 041

$$E(w) = \frac{1}{2}\sum_{i=1}^{n}\left(z_i - w^T x_i\right)^2 + \frac{\lambda}{2}\|w\|^2$$

$$\nabla E(w) = 0 \Leftrightarrow \nabla\left(\frac{1}{2}\cdot(z - Xw)^T(z - X.w) + \frac{\lambda}{2}w^T w\right) = 0 \Leftrightarrow$$

$$\Leftrightarrow -2X^T z + 2X^T \cdot X \cdot w + 2\lambda \cdot w = 0 \Leftrightarrow$$

$$\Leftrightarrow X^T z = \left(X^T \cdot X + \lambda I\right) w \Leftrightarrow$$

$$\Rightarrow w = X^T z \times \left(X^T \cdot X + \lambda I\right)^{-1} = \left(\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{pmatrix} \times \right.$$

$$= \left(\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{pmatrix} \times \begin{pmatrix} 1 & 0{,}8 & 0{,}64 & 0{,}512 \\ 1 & 1 & 1 & 1 \\ 1 & 1{,}2 & 1{,}44 & 1{,}728 \\ 1 & 1{,}4 & 1{,}96 & 2{,}744 \\ 1 & 1{,}6 & 2{,}56 & 4{,}096 \end{pmatrix} + 2 \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right)^{-1} \times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{pmatrix} \times \begin{pmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{pmatrix} =$$

$$= \left(\begin{bmatrix} 5 & 6 & 7{,}6 & 10{,}08 \\ 6 & 7{,}6 & 10{,}08 & 13{,}88 \\ 7{,}6 & 10{,}08 & 13{,}88 & 19{,}68 \\ 10{,}8 & 13{,}88 & 19{,}68 & 28{,}56 \end{bmatrix} + \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}\right)^{-1} \times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{pmatrix} \times \begin{pmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{pmatrix} =$$

$$= \begin{bmatrix} 7 & 6 & 7{,}6 & 10{,}08 \\ 6 & 9{,}6 & 10{,}0 & 13{,}88 \\ 7{,}6 & 10{,}08 & 15{,}8 & 19{,}68 \\ 10{,}08 & 13{,}88 & 19{,}6 & 30{,}56 \end{bmatrix}^{-1} \times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{pmatrix} \times \begin{pmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{pmatrix} =$$

Aprendizagem 2021/22

Homework III – Group 041

$$= \begin{pmatrix} 7,0\ 1 \\ 4,165\ 41 \\ 1,1861 \\ -1,3 \end{pmatrix}$$

$$y(x,w) = 7,091 + 4,6541\,x - 1,3615\,x^2 - 1,3\,x^3$$

2)

$$y(x,w) = 7,091 + 4,6541\,x - 1,3615\,x^2 - 1,3\,x^3$$

$$y(x_1,w) = 7,091 + 4,6541 \times 0,8 - 1,3615 \times (0,8)^2 - 1,3 \times (0,8)^3$$
$$= 8,8933$$

$$y(x_2,w) = 7,091 + 4,6541 \times 1 - 1,3615 \times 1 - 1,3 \times 1 = 8,484$$

$$y(x_3,w) = 7,091 + 4,6541 \times 1,2 - 1,3615 \times (1,2)^2 - 1,3 \times (1,2)^3 =$$
$$= 7,605$$

$$y(x_4,w) = 7,091 + 4,6541 \times 1,4 - 1,3615 \times (1,4)^2 - 1,3 \times (1,4)^3 =$$
$$= 6,1848$$

$$(24 - 8,8933)^2 = 228,2124$$

$$(20 - 8,484)^2 = 132,61$$

$$(10 - 7,605)^2 = 5,7360$$

$$(13 - 6,1948)^2 = 46,3094$$

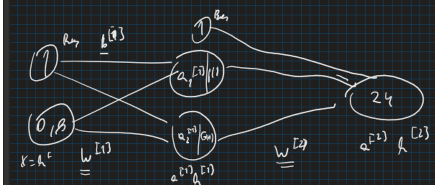
$$(12 - 4,1912)^2 = 60,977$$

$$\sqrt{\frac{\sum_{i=1}^{5} (y_i - \hat{y}_i)^2}{5}}$$

$$\sqrt{\frac{\sum_{i=1}^{5}(y_i-\hat{y}_i)^2}{5}} = \sqrt{\frac{228,7724+132,61+5,7360+46,3054+60,5774}{5}} = $$

$$= 9,7349$$

3)



Forward Propagation

$$\underline{a}^{[1]} = \underline{\underline{W}}^{[1]}\underline{x} + \underline{b}^{[1]} \qquad \underline{h}^{[1]} = f(\underline{a}^{[1]}) = \begin{bmatrix} f(a_1^{[1]}) \\ f(a_2^{[1]}) \\ f(a_3^{[1]}) \end{bmatrix}$$

$$\underline{a}^{[2]} = \underline{\underline{W}}^{[2]}\underline{h}^{[1]} + \underline{b}^{[2]}$$

$$h^{[2]} = f(\underline{a}^{[2]}) = \hat{z}$$

$$f(x) = e^{0,1x}$$

$$\frac{\partial f(x)}{\partial x} = 0,1 \times e^{0,1x}$$

half squared error loss $= \frac{1}{2N}\sum_{i=1}^{N}(x-\hat{x})^2$

$\underline{x} \in \mathbb{R}^1 \quad \underline{a}^{[1]}, \underline{h}^{[1]} \in \mathbb{R}^2$

$\underline{\underline{W}}^{[1]} \in \mathbb{R}^{2\times 1} \longrightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} \times [0] = 0$

$\underline{b}^{[1]} \in \mathbb{R}^{2\times 1} \quad b^{[2]} \in \mathbb{R}^1$

$a^{[2]}, h^{[2]} \in \mathbb{R}^1$

$\underline{\underline{W}}^{[2]} \in \mathbb{R}^{1\times 2} \longrightarrow [0 \quad 0] \times \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$

$$\frac{\partial E(z,\hat{z})}{\partial \underline{\underline{W}}^{[1]}} \ ?$$

$$\frac{\partial E(z,\hat{z})}{\partial \underline{b}^{[1]}} = ?$$

$$\frac{\partial E(z,\hat{z})}{\partial w}$$

$$\underline{w} = \underline{w}^{old} - \eta\frac{\partial E}{\partial w}$$

Inicialização:

$$\underline{\underline{W}}^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\underline{\underline{W}}^{[2]} = [1\ 1]$$

$$\underline{b}^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b^{[2]} = 1$$

half squared error $\longrightarrow E(w) = \frac{1}{2}\sum_{k=1}^{N} t^{(k)} - \text{Output}(x^{(k)}; w)$

1º passo forward propagation

$$a^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times 0,8 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,8 \\ 0,8 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1,8 \\ 1,8 \end{bmatrix}$$

$$\underline{h}^{[1]} = f(\underline{a}^{[1]}) \approx f\left(\begin{bmatrix} 1,8 \\ 1,8 \end{bmatrix}\right) = \begin{bmatrix} 1,197 \\ 1,197 \end{bmatrix}$$

$$a^{[2]} = [1\ 1] \times \begin{bmatrix} 1,197 \\ 1,197 \end{bmatrix} + 1 = 3,394$$

$$h^{[2]} = f(a^{[2]}) = f(3,394) = e^{0,1\times 3,394} = 1,404$$

Aprendizagem 2021/22

Homework III – Group 041

2º passo: Back propagation

$$\frac{\partial E}{\partial W^{[2]}} \overset{①}{=} \frac{\partial E}{\partial a^{[2]}} \circ h^{[1]T} \overset{④}{=} \underbrace{\frac{\partial E}{\partial h^{[2]}} \cdot f'(a^{[2]})}_{} \circ h^{[1]T} \overset{⑤}{=} \frac{\partial E}{\partial h^{[2]}} \cdot \left(+0,1\, e^{+0,1\times 3,394}\right) \circ [1.197, 1.197] =$$

$$= (h^{[2]} - z) \circ (0,1404) \circ [1,197, 1,197] =$$

$$= (1,404 - 24) \cdot (0,1404) \cdot [1.197, 1.197] =$$

$$= -3,1725 \times [1.197, 1.197] =$$

$$= [-3,7975, -3,7975]$$

⑤

$$\frac{\partial E}{\partial \hat z} = \frac{1}{2} \times 2 \times (z - \hat z) \times (-1) = (\hat z - z)$$

$$\frac{\partial E}{\partial h^{[2]}} = \frac{\partial E}{\partial \hat z} = \frac{\partial}{\partial \hat z}\left[\frac{1}{2\times 1}(z - \hat z)^2\right]$$

$$\frac{\partial E}{\partial b^{[2]}} = \frac{\partial E}{\partial a^{[2]}} \overset{③}{=} -3,1725$$

$$h^{[2]} = f(a^{[2]}) \Rightarrow \frac{\partial E}{\partial a^{[2]}} \overset{④}{=} \underbrace{\frac{\partial E}{\partial h^{[2]}} \times f'(a^{[2]})}_{}$$

$$(v_1\ v_2) \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = v_1 w_1 + v_2 w_2$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 w_1 \\ v_2 w_2 \end{bmatrix}$$

Chain rule

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u} \times \frac{\partial u}{\partial x}$$

$$\frac{\partial E}{\partial C}$$

$C = AB$:

$$\frac{\partial E}{\partial A} \overset{①}{=} \frac{\partial E}{\partial C} \times B^T$$

$$\frac{\partial E}{\partial B} \overset{②}{=} A^T \frac{\partial E}{\partial C}$$

$C = A + B$:

$$\frac{\partial E}{\partial A} = \frac{\partial E}{\partial C} \times \cancel{\frac{\partial C}{\partial A}}^{\,1}$$

$$\frac{\partial E}{\partial B} = \frac{\partial E}{\partial C} \times \cancel{\frac{\partial C}{\partial B}}^{\,1} \quad ③$$

$$C = f(A) \rightarrow \frac{\partial E}{\partial A} = \frac{\partial E}{\partial C} \circ f'(A) \quad ④$$

Forward Propagation

$$a^{[1]} = W^{[1]} x + b^{[1]} \qquad h^{[1]} = f(a^{[1]}) = \begin{bmatrix} f(a_1^{[1]}) \\ f(a_2^{[1]}) \\ f(a_3^{[1]}) \end{bmatrix}$$

$$a^{[2]} = W^{[2]} h^{[1]} + b^{[2]}$$

$$h^{[2]} = f(a^{[2]}) = \hat z$$

⑤

$$\frac{\partial E}{\partial \hat z} = \frac{1}{2} \times 2 \times (z - \hat z) \times (-1) = (\hat z - z)$$

$$\frac{\partial E}{\partial h^{[2]}} = \frac{\partial E}{\partial \hat z} = \frac{\partial}{\partial \hat z}\left[\frac{1}{2\times 1}(z - \hat z)^2\right]$$

$$\frac{\partial E}{\partial b^{[2]}} = \frac{\partial E}{\partial a^{[2]}} \overset{③}{=} -3,1725$$

$$h^{[2]} = f(a^{[2]}) \Rightarrow \frac{\partial E}{\partial a^{[2]}} \overset{④}{=} \underbrace{\frac{\partial E}{\partial h^{[2]}} \times f'(a^{[2]})}_{}$$

$$\frac{\partial E}{\partial W^{[1]}} = \frac{\partial E}{\partial a^{[1]}} \times x^T = \frac{\partial E}{\partial a^{[1]}} \cdot f'(a^{[1]}$$

$$= ((h^{[2]} - z) \circ h'$$

TÉCNICO LISBOA

Aprendizagem 2021/22

Homework III – Group 041

$$\frac{\partial E}{\partial W^{[1]}} = \frac{\partial E}{\partial a^{[1]}} \times X^T = \frac{\partial E}{\partial a^{[1]}} \circ f'\left(a^{[1]}\right) \circ X^T =$$

$$= \left(\left(a^{[1]} - z\right) \circ f'\left(\begin{bmatrix} 4,8 \\ 4,8 \end{bmatrix}\right)\right) \circ 0,8$$

$$\overset{⑥}{=} \left(W^{[2]T} \frac{\partial E}{\partial a^{[2]}} \cdot \begin{bmatrix} 0,1137 \\ 0,1137 \end{bmatrix}\right) \cdot 0,8$$

$$= \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \times (-3,1725) \times \begin{bmatrix} 0,1187 \\ 0,1137 \end{bmatrix}\right) \cdot 0,8$$

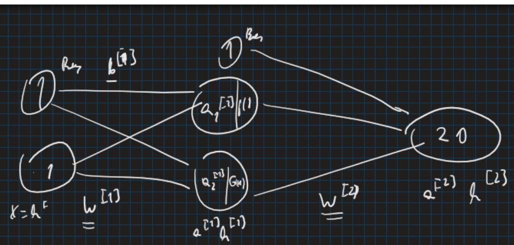
$$= \left(\begin{bmatrix} -3,1725 \\ -3,1725 \end{bmatrix} \times \begin{bmatrix} 0,1157 \\ 0,1187 \end{bmatrix}\right) \times 0,8$$

$$= \underbrace{\begin{pmatrix} -0,3797 \\ -0,3797 \end{pmatrix}}_{\frac{\partial E}{\partial a^{[2]}}} \times 0,8 = \begin{bmatrix} -0,30376 \\ -0,30376 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[2]}} \overset{③}{=} \frac{\partial E}{\partial a^{[2]}} = \begin{bmatrix} -0,3797 \\ -0,3797 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 041



$x \in \mathbb{R}^1 \quad e^{[1]}, h^{[1]} \in \mathbb{R}^2$

$W^{[1]} \in \mathbb{R}^{2\times 1} \longrightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} \times [0] = 0$

$b^{[1]} \in \mathbb{R}^{2\times 1} \quad b^{[2]} \in \mathbb{R}^1$

$e^{[2]}, h^{[2]} \in \mathbb{R}^1$

$W^{[2]} \in \mathbb{R}^{1\times 2} \longrightarrow [0 \ \ 0] \times \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$

Inicialização:

$W^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$W^{[2]} = [1 \ 1]$

$b^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b^{[2]} = 1$

Forward Propagation

$e^{[1]} = W^{[1]} x + b^{[1]} \qquad h^{[1]} = f(e^{[1]}) = \begin{bmatrix} f(e_1^{[1]}) \\ f(e_2^{[1]}) \\ f(e_3^{[1]}) \end{bmatrix}$

$e^{[2]} = W^{[2]} h^{[1]} + b^{[2]}$

$h^{[2]} = f(e^{[2]}) = \hat{z}$

1º passo forward propagation

$e^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times 1 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$h^{[1]} = f(e^{[1]}) \approx f\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1{,}2214 \\ 1{,}2214 \end{bmatrix}$

$e^{[2]} = [1 \ 1] \times \begin{bmatrix} 1{,}2214 \\ 1{,}2214 \end{bmatrix} + 1 = 3{,}4428$

$h^{[2]} = f(e^{[2]}) = f(3{,}4428) = e^{0{,}1 \times 3{,}4428} = 1{,}411$

2º passo: Back propagation

$\frac{\partial E}{\partial W^{[2]}} \overset{1}{=} \frac{\partial E}{\partial e^{[2]}} \cdot h^{[1]T} \overset{2}{=} \frac{\partial E}{\partial h^{[2]}} \cdot f'(e^{[2]}) \cdot h^{[1]T} = \overset{3}{\frac{\partial E}{\partial h^{[2]}}} \cdot \left(0{,}1\, e^{0{,}1\times 3{,}4428}\right) \cdot [1.2214, 1.2214] =$

$= (h^{[2]} - z) \cdot (0{,}1411) \cdot [1{,}2214, 1{,}2214]$

$= (1{,}411 - 20) \cdot (0{,}1411) \cdot [1.2214, 1.2214]$

$= -2{,}6229 \times [1.2214, 1.2214] =$

$= [-3{,}2036, -3{,}2036]$

3

$\frac{\partial E}{\partial \hat{z}} = \frac{1}{2} \times 2 \times (z - \hat{z}) \times (-1) = (\hat{z} - z)$

$\frac{\partial E}{\partial h^{[2]}} = \frac{\partial E}{\partial \hat{z}} = \frac{\partial}{\partial \hat{z}} \left[\frac{1}{2}(z - \hat{z})^2\right]$

$\frac{\partial E}{\partial e^{[2]}} = -2{,}6229$

logo

$\frac{\partial E}{\partial b^{[2]}} \overset{4}{=} \frac{\partial E}{\partial e^{[2]}} = -2{,}6229$

$[v_1 \ v_2] \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = v_1 w_1 + v_2 w_2$

$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 w_1 \\ v_2 w_2 \end{bmatrix}$

Aprendizagem 2021/22

Homework III – Group 041

Chain rule

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u} \times \frac{\partial u}{\partial x}$$

$$\frac{\partial E}{\partial C}$$

$$\underline{C} = \underline{A}\,\underline{B} \quad \frac{\partial E}{\partial A} \overset{1}{=} \frac{\partial E}{\partial C} \times B^T \qquad \frac{\partial E}{\partial B} \overset{2}{=} A^T \frac{\partial E}{\partial C}$$

$$\underline{C} = A + B \quad \frac{\partial E}{\partial A} = \frac{\partial E}{\partial C} \times \cancel{\frac{\partial C}{\partial A}}^{\,1} \qquad \frac{\partial E}{\partial B} = \frac{\partial E}{\partial C} \times \cancel{\frac{\partial C}{\partial B}}^{\,1} \quad 3$$

$$\underline{C} = f(A) \rightarrow \frac{\partial E}{\partial A} = \frac{\partial E}{\partial C} \circ f'(A) \quad 4$$

Foward Propagation

$$\underline{e}^{[1]} = \underline{\underline{W}}^{[1]}\,\underline{x} + b^{[1]} \qquad \underline{h}^{[1]} = f(\underline{e}^{[1]}) = \begin{bmatrix} f(e_1^{[1]}) \\ f(e_2^{[1]}) \\ f(e_3^{[1]}) \end{bmatrix}$$

$$\underline{e}^{[2]} = W^{[2]}\,\underline{h}^{[1]} + \underline{b}^{[2]}$$

$$h^{[2]} = (\underline{e}^{[2]}) = \hat{\underline{z}}$$

5

$$\frac{\partial E}{\partial \hat{z}} = \frac{1}{2} \times 2 \times (z - \hat{z}) \times (-1) = (\hat{z} - z)$$

$$\frac{\partial E}{\partial h^{[2]}} = \frac{\partial E}{\partial \hat{z}} = \frac{\partial}{\partial \hat{z}}\left[\frac{1}{2 \times 1}(z - \hat{z})^2\right]$$

$$= (h^{[2]} - z) \circ (0{,}1411) \cdot [1{,}2214, 1\ldots$$

$$= (1{,}411 - 20) \cdot (0{,}1411) \cdot [1{,}2214, 1\ldots$$

$$= -2{,}6229 \times [1{,}2214, 1{,}2214] =$$

$$= [-3{,}2036, -3{,}2036]$$

$$\frac{\partial E}{\partial a^{[2]}} = -2{,}6229$$

$$(v_1\ v_2) \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = v_1 w_1 \ldots$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 w_1 \\ v_2 w_2 \end{bmatrix}$$

logo

$$\frac{\partial E}{\partial b^{[2]}} \overset{3}{=} \frac{\partial E}{\partial a^{[2]}} = -2{,}6229$$

$$\frac{\partial E}{\partial W^{[1]}} = \frac{\partial E}{\partial a^{[1]}} \times X^T = \frac{\partial E}{\partial a^{[1]}} \circ f'(a^{[1]}) \cdot X^T =$$

$$= \left((a^{[1]} - z) \circ f'\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)\right) \cdot 1$$

$$\frac{\partial E}{\partial W^{[1]}} = \frac{\partial E}{\partial a^{[1]}} \times X^T = \frac{\partial E}{\partial a^{[1]}} \circ f'(a^{[1]}) \cdot X^T =$$

$$= \left((a^{[1]} - z) \circ f'\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)\right) \cdot 1$$

$$\overset{6}{=} \left(W^{[2]T} \frac{\partial E}{\partial a^{[2]}} \cdot \begin{bmatrix} 0{,}1221 \\ 0{,}1221 \end{bmatrix}\right) \cdot 1$$

$$= \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \times (-2{,}6229) \times \begin{bmatrix} 0{,}1211 \\ 0{,}1221 \end{bmatrix}\right) \cdot 1$$

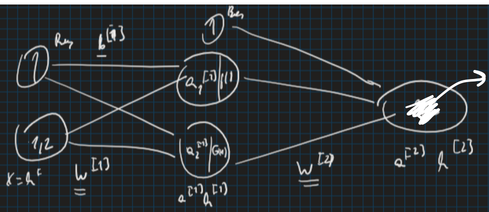
$$= \left(\begin{bmatrix} -2{,}6229 \\ -2{,}6229 \end{bmatrix} \times \begin{bmatrix} 0{,}1211 \\ 0{,}1211 \end{bmatrix}\right) \times 1$$

$$= \underbrace{\begin{pmatrix} -0{,}3176 \\ -0{,}3176 \end{pmatrix}}_{\frac{\partial E}{\partial a^{[1]}}} \times 1 = \begin{bmatrix} -0{,}3176 \\ -0{,}3176 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[1]}} \overset{3}{=} \frac{\partial E}{\partial a^{[1]}} = \begin{bmatrix} -0{,}3176 \\ -0{,}3176 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 041



Inicialização:

$W^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$W^{[2]} = [1\ 1]$

$b^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $b^{[2]} = 1$

$x \in \mathbb{R}^1$ $a^{[1]}, h^{[1]} \in \mathbb{R}^2$

$W^{[1]} \in \mathbb{R}^{2\times 1} \longrightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} \times [0] = 0$

$b^{[1]} \in \mathbb{R}^{2\times 1}$ $b^{[2]} \in \mathbb{R}^1$

$a^{[2]}, h^{[2]} \in \mathbb{R}^1$

$W^{[2]} \in \mathbb{R}^{1\times 2} \longrightarrow [0\ \ 0] \times \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$

$\frac{\partial}{\partial b^{[1]}} = \frac{\partial}{\partial a^{[1]}} =$

Forward Propagation

$a^{[1]} = W^{[1]} x + b^{[1]}$ $h^{[1]} = f(a^{[1]}) = \begin{bmatrix} f(a_1^{[1]}) \\ f(a_2^{[1]}) \\ f(a_3^{[1]}) \end{bmatrix}$

$a^{[2]} = W^{[2]} h^{[1]} + b^{[2]}$

$h^{[2]} = f(a^{[2]}) = \hat{z}$

1º passo forward propagation

$a^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times 1{,}2 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1{,}2 \\ 1{,}2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2{,}2 \\ 2{,}2 \end{bmatrix}$

$h^{[1]} = f(a^{[1]}) \approx f\left(\begin{bmatrix} 2{,}2 \\ 2{,}2 \end{bmatrix}\right) = \begin{bmatrix} 1{,}2461 \\ 1{,}2461 \end{bmatrix}$

$a^{[2]} = [1\ 1] \times \begin{bmatrix} 1{,}2461 \\ 1{,}2461 \end{bmatrix} + 1 = 3{,}4922$

$h^{[2]} = f(a^{[2]}) = f(3{,}4922) = e^{0{,}1 \times 3{,}4922} = 1{,}418$

2º passo: Back propagation

$\frac{\partial E}{\partial W^{[2]}} \overset{1}{=} \frac{\partial E}{\partial a^{[2]}} \cdot h^{[1]T} \overset{2}{=} \frac{\partial E}{\partial h^{[2]}} \cdot f'(a^{[2]}) \cdot h^{[1]T} = \overset{3}{\frac{\partial E}{\partial h^{[2]}}} \cdot \left(+0{,}1\, e^{+0{,}1\times 3{,}4922}\right) \cdot [1.2461,\ 1.2461]$

$= (h^{[2]} - z) \cdot (0{,}1418) \cdot [1{,}2461,\ 1{,}2461]$

$= (1{,}418 - 10) \cdot (0{,}1418) \cdot [1.2461,\ 1.2461]$

$= -1{,}2169 \times [1.2461,\ 1.2461] =$

$= [-1{,}5164,\ -1{,}5164]$

(3)

$\frac{\partial E}{\partial \hat{z}} = \frac{1}{2} \times 2 \times (z - \hat{z}) \times (-1) = (\hat{z} - z)$

$\frac{\partial E}{\partial h^{[2]}} = \frac{\partial E}{\partial \hat{z}} = \frac{\partial}{\partial \hat{z}} \left[\frac{1}{2}(z - \hat{z})^2\right]$

$\frac{\partial E}{\partial a^{[2]}} = -1{,}2169$

logo

$[v_1\ v_2] \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = v_1 w_1 + v_2$

$$\frac{\partial E}{\partial b^{[2]}} \overset{(3)}{=} \frac{\partial E}{\partial a^{[2]}} = -1{,}2169 \qquad \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 w_1 \\ v_2 w_2 \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{[1]}} = \frac{\partial E}{\partial a^{[1]}} \times X^T \overset{(4)}{=} \frac{\partial E}{\partial a^{[1]}} \circ f'(a^{[1]}) \circ X^T =$$

$$= \left((a^{[2]} - t) \circ f'\left(\begin{bmatrix} 2{,}2 \\ 2{,}2 \end{bmatrix} \right) \right) \circ 1{,}2$$

$$\overset{(6)}{=} \left(W^{[2]T} \frac{\partial E}{\partial a^{[2]}} \circ \begin{bmatrix} 0{,}1246 \\ 0{,}1246 \end{bmatrix} \right) \circ 1{,}2$$

$$= \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \times (-1{,}2169) \times \begin{bmatrix} 0{,}1246 \\ 0{,}1246 \end{bmatrix} \right) \circ 1{,}2$$

$$= \left(\begin{bmatrix} -1{,}2169 \\ -1{,}2169 \end{bmatrix} \times \begin{bmatrix} 0{,}1246 \\ 0{,}1246 \end{bmatrix} \right) \times 1.2$$

$$= \underbrace{\begin{bmatrix} -0{,}1516 \\ -0{,}1516 \end{bmatrix}}_{\frac{\partial E}{\partial a^{[1]}}} \times 1.2 = \begin{bmatrix} -0{,}182 \\ -0{,}182 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[1]}} \overset{(3)}{=} \frac{\partial E}{\partial a^{[1]}} = \begin{bmatrix} -0{,}1516 \\ -0{,}1516 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 041

Por fim:

$$\frac{\partial E}{\partial b^{[1]}}(\text{total}) = \frac{\partial E}{\partial a^{[1]}}(\text{total}) = \begin{bmatrix} -0{,}3787 \\ -0{,}3787 \end{bmatrix} + \begin{bmatrix} -0{,}3176 \\ -0{,}3176 \end{bmatrix} + \begin{bmatrix} -0{,}1526 \\ -0{,}1516 \end{bmatrix} = \begin{bmatrix} -0{,}8489 \\ -0{,}8489 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[2]}}(\text{total}) = \frac{\partial E}{\partial a^{[2]}}(\text{total}) = -3{,}1725 + (-2{,}6225) + (-1{,}2169) = -7{,}0123$$

$$\frac{\partial E}{\partial W^{[2]}}(\text{total}) = [-3{,}17875, -3{,}17875] + [-3{,}2036, -3{,}2036] + [-1{,}5164, -1{,}5164] = [-8.518, -8.518]$$

$$\frac{\partial E}{\partial W^{[1]}}(\text{total}) = \begin{bmatrix} -0{,}30376 \\ -0{,}30376 \end{bmatrix} + \begin{bmatrix} -0{,}3176 \\ -0{,}3176 \end{bmatrix} + \begin{bmatrix} -0{,}182 \\ -0{,}182 \end{bmatrix} = \begin{bmatrix} -0{,}8034 \\ -0{,}8034 \end{bmatrix}$$

$$b^{[1]\,new} = b^{[1]\,old} - \eta \frac{\partial E}{\partial b^{[1]}}^{total} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}8489 \\ -0{,}8489 \end{bmatrix} = \begin{bmatrix} 1{,}0872 \\ 1{,}0872 \end{bmatrix}$$

$$b^{[2]\,new} = b^{[2]\,old} - \eta \frac{\partial E}{\partial b^{[2]}} = 1 - 0{,}1 \times (-7{,}0123) = 1{,}71533$$

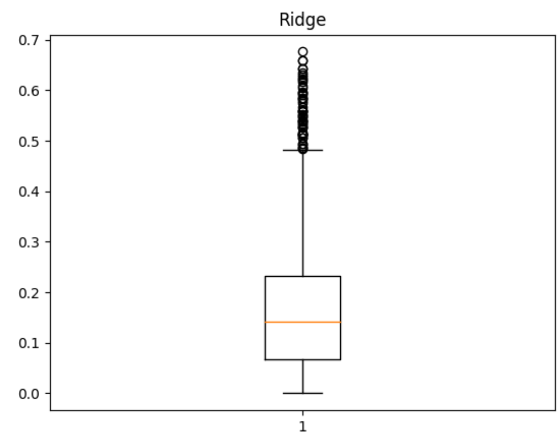
$$W^{[1]\,new} = W^{[1]\,old} - \eta \frac{\partial E}{\partial W^{[1]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \times \begin{bmatrix} -0{,}8034 \\ -0{,}8034 \end{bmatrix} = \begin{bmatrix} 1{,}080 \\ 1{,}080 \end{bmatrix}$$

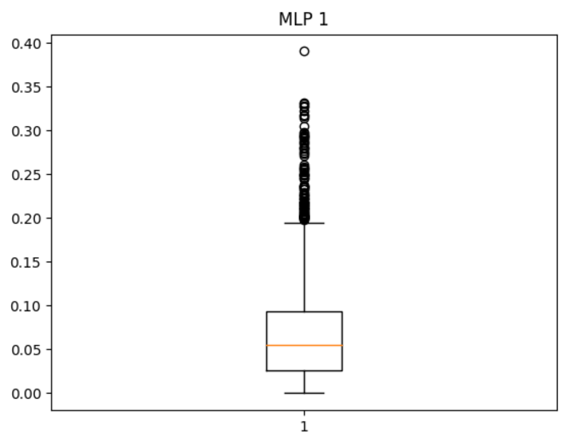
$$W^{[2]\,new} = W^{[2]\,old} - \eta \frac{\partial E}{\partial W^{[2]}} = [1 \ 1] - 0{,}1 \times [-8.518, -8.518] = \begin{bmatrix} 1{,}8518 \\ 1{,}8518 \end{bmatrix}$$

II. Programming and critical analysis

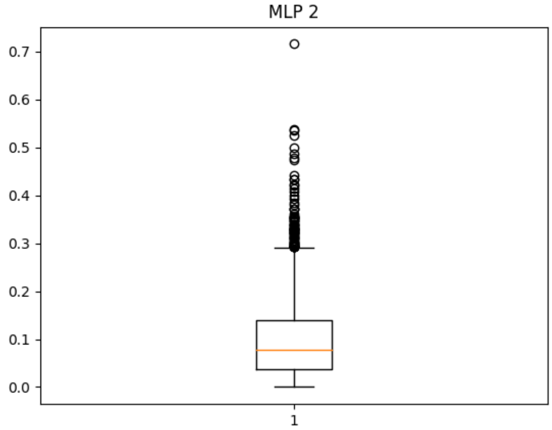
4) Ridge MAE: 0.162829976437694
MLP 1 MAE: 0.0680414073796843
MLP 2 MAE: 0.0978071820387748

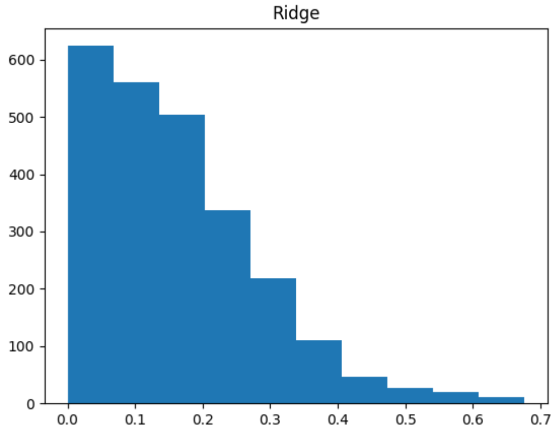
5)

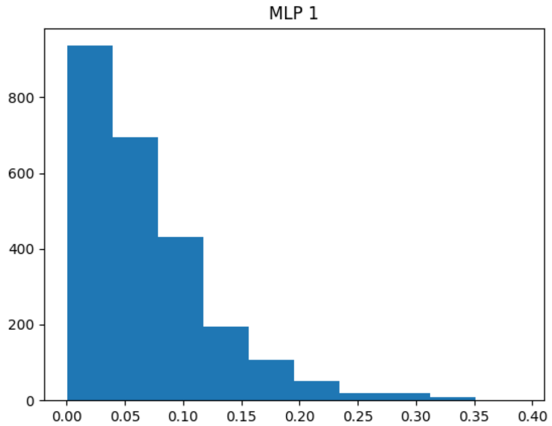


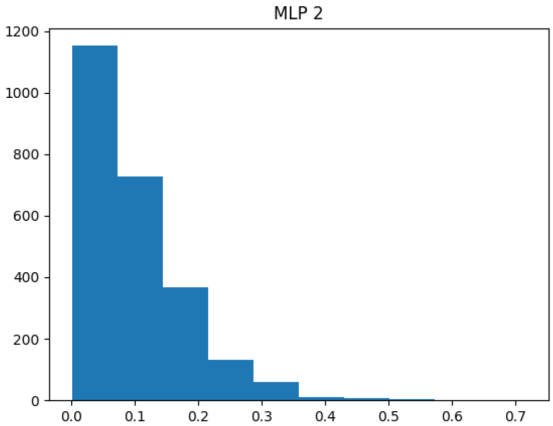


Aprendizagem 2021/22
Homework III – Group 041









6) Iterações do MLP 1: 452
Iterações do MLP 2: 77

7) As iterações do MLP 1 são maiores que as do MLP 2 o que é o contrário do esperado. Tem-se que quando o modelo aprende com early restart, como é o caso do MLP 1, o processo de aprendizagem é parado apenas quando não é detetado um melhoramento na pontuação de validação igual ou superior a um determinado valor durante um determinado número de iterações. Isto é feito para contornar o potencial barulho estatístico no dataset. Tendo isto em consideração, a presença de bastante barulho estatístico no dataset pode motivar o maior número de iterações em MLP 1, uma vez que a paragem do processo é adiada várias vezes.
Quanto à diferença de performance, o MLP 1 resultou num modelo que não está over-fitted, o que resulta numa melhor performance comparativamente ao MLP 2.

III. APPENDIX

Paste your programming code here using Consolas 9pt or 10pt.

Use **highlighting** or colored text to facilitate the analysis by your faculty hosts.

```
import pandas as pd
import numpy as np

#*##################################################################################
```

```
#* 4)
#*###################################################################################

#* import data
from scipy.io.arff import loadarff
data = loadarff("kin8nm.arff")
df = pd.DataFrame(data[0])
num_columns = df.shape[1]

#* partition data
from sklearn.model_selection import train_test_split
X, y = df.iloc[:, 0:num_columns-1], df['y']
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.70, random_state=0)

#* linear regressor with Ridge regularization
from sklearn.linear_model import Ridge
rr = Ridge(alpha=0.1)

#* MLP_1 and MLP_2 regressors
from sklearn.neural_network import MLPRegressor
MLP1r = MLPRegressor(hidden_layer_sizes=(10, 10), activation="tanh", max_iter=500, random_state=0,
early_stopping=True)
MLP2r = MLPRegressor(hidden_layer_sizes=(10, 10), activation="tanh", max_iter=500, random_state=0,
early_stopping=False)

#* learn (".values" was added to avoid warnings)
rr.fit(X_train.values, y_train.values)
MLP1r.fit(X_train.values, y_train.values)
MLP2r.fit(X_train.values, y_train.values)

#* predict (".values" was added to avoid warnings)
rr_y_pred = rr.predict(X_test.values)
MLP1r_y_pred = MLP1r.predict(X_test.values)
MLP2r_y_pred = MLP2r.predict(X_test.values)

#* compute MAE
from sklearn.metrics import mean_absolute_error
y_true = y_test
rr_MAE = mean_absolute_error(y_true, rr_y_pred)
MLP1r_MAE = mean_absolute_error(y_true, MLP1r_y_pred)
MLP2r_MAE = mean_absolute_error(y_true, MLP2r_y_pred)
print("Ridge MAE: " + str(rr_MAE))
print("MLP1 MAE: " + str(MLP1r_MAE))
print("MLP2 MAE: " + str(MLP2r_MAE))

#*###################################################################################
#* 5)
```

```
#*#####################################################################################

#* array of residuals (in absolute value)
rr_residuals = abs(y_true - rr_y_pred)
MLP1r_residuals = abs(y_true - MLP1r_y_pred)
MLP2r_residuals = abs(y_true - MLP2r_y_pred)

#* plot
import matplotlib.pyplot as plt
plt.boxplot(x=rr_residuals)
plt.title(label="Ridge")
plt.show()
plt.boxplot(x=MLP1r_residuals)
plt.title(label="MLP 1")
plt.show()
plt.boxplot(x=MLP2r_residuals)
plt.title(label="MLP 2")
plt.show()
plt.hist(x=rr_residuals)
plt.title(label="Ridge")
plt.show()
plt.hist(x=MLP1r_residuals)
plt.title(label="MLP 1")
plt.show()
plt.hist(x=MLP2r_residuals)
plt.title(label="MLP 2")
plt.show()

#*#####################################################################################
#* 6)
#*#####################################################################################

MLP1r_iterations = MLP1r.n_iter_
MLP2r_iterations = MLP2r.n_iter_
print('MLP1 Iterations: ' + str(MLP1r_iterations))
print('MLP2 Iterations: ' + str(MLP2r_iterations))
```

END