Aprendizagem 2022/23

Homework III – Group 107

(ist1103811, ist1103479)

I. Pen-and-paper

1)

1) $x = \left\{ \begin{pmatrix} 0.7 \\ -0.3 \end{pmatrix}, \begin{pmatrix} 0.4 \\ 0.5 \end{pmatrix}, \begin{pmatrix} -0.2 \\ 0.8 \end{pmatrix}, \begin{pmatrix} -0.4 \\ 0.3 \end{pmatrix} \right\}$, targets $= (0.8, 0.6, 0.3, 0.3)$

$$\phi_j(x) = e^{-\frac{\|x - c_j\|^2}{2}}$$

$$c_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad c_3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$\phi_1(x_1) = e^{-\frac{0.7^2 + 0.3^2}{2}} = 0.74826$ $\qquad \phi_1(x_2) = 0.81465$ $\qquad \phi_1(x_3) = 0.71177$

$\phi_2(x_1) = e^{-\frac{(0.7-1)^2 + (-0.3+1)^2}{2}} = 0.74826$ $\qquad \phi_2(x_2) = 0.27117$ $\qquad \phi_2(x_3) = 0.09633$

$\phi_3(x_1) = e^{-\frac{(0.7+1)^2 + (-0.3-1)^2}{2}} = 0.10127$ $\qquad \phi_3(x_2) = 0.33121$ $\qquad \phi_3(x_3) = 0.71177$

$\phi_1(x_4) = 0.8825$ $\qquad \phi_2(x_4) = 0.16122$ $\qquad \phi_3(x_4) = 0.65377$

(Continua na próxima página)

TÉCNICO LISBOA

Aprendizagem 2022/23

Homework III – Group 107

(ist1103811, ist1103479)

$$\begin{array}{c} \\ \phi(x) = \end{array} \begin{array}{ccc} c_1 & c_2 & c_3 \\ \end{array}$$

$$\phi(x) = \begin{pmatrix} 0.74826 & 0.74826 & 0.10127 \\ 0.81465 & 0.27117 & 0.33121 \\ 0.71177 & 0.09633 & 0.71177 \\ 0.8825 & 0.16122 & 0.65377 \end{pmatrix} \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}$$

$$\rightarrow X = \begin{pmatrix} 1 & 0.74826 & 0.74826 & 0.10127 \\ 1 & 0.81465 & 0.27117 & 0.33121 \\ 1 & 0.71177 & 0.09633 & 0.71177 \\ 1 & 0.8825 & 0.16122 & 0.65377 \end{pmatrix}$$

$$\text{Ridge} \rightarrow (X^T X + 0.1\,I)^{-1} X^T z =$$

$$= \left(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0.74826 & 0.81465 & 0.71177 & 0.8825 \\ 0.74826 & 0.27117 & 0.09633 & 0.16122 \\ 0.10127 & 0.33121 & 0.71177 & 0.65377 \end{pmatrix} \begin{pmatrix} 1 & 0.74826 & 0.74826 & 0.10127 \\ 1 & 0.81465 & 0.27117 & 0.33121 \\ 1 & 0.71177 & 0.09633 & 0.71177 \\ 1 & 0.8825 & 0.16122 & 0.65377 \end{pmatrix} + 0.1 I \right)^{-1} X^T z =$$

$$= \left(\begin{pmatrix} 4 & 3.15718 & 1.27698 & 1.79802 \\ 3.15718 & 2.50897 & 0.99165 & 1.42916 \\ 1.27698 & 0.99165 & 0.6687 & 0.33955 \\ 1.79802 & 1.42916 & 0.33955 & 1.05399 \end{pmatrix} + \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix} \right)^{-1} X^T z$$

$$= \begin{pmatrix} 4.54826 & -3.77682 & -1.86117 & -1.86155 \\ -3.77682 & 5.98285 & -0.88543 & -1.26432 \\ -1.86117 & -0.88543 & 4.33276 & 2.72156 \\ -1.86155 & -1.26432 & 2.72156 & 4.53204 \end{pmatrix} X^T z =$$

$$= \begin{pmatrix} 0.33914 \\ 0.19945 \\ 0.40096 \\ -0.296 \end{pmatrix} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

b)

	z	$\hat{z}$	$z - \hat{z}$
x_1	0.8	0.75472	0.04528
x_2	0.6	0.50828	0.09172
x_3	0.3	0.30552	-0.00552
x_4	0.3	0.38191	-0.08191

C4

$$\hat{z}(x) = w_0 + w_1\phi_1(x) + w_2\phi_2(x) + w_3\phi_3(x)$$

$$\hat{z}(x_1) = 0.33914 + 0.19945 \times 0.74826 + 0.40096 \times 0.74826 - 0.296 \times 0.10127 = 0.75472$$

$$\hat{z}(x_2) = 0.50828$$

$$\hat{z}(x_3) = 0.30552$$

$$\hat{z}(x_4) = 0.38191$$

$$RMSE = \sqrt{\frac{1}{4}\sum_{1}^{4}(z_i - \hat{z}_i)^2} = \frac{1}{2}\sqrt{0.04528^2 + 0.09172^2 + (-0.00552)^2 + (-0.08191)^2} =$$

$$= \frac{1}{2} \times \sqrt{0.0172} = \frac{1}{2} \times 0.13116 = 0.06558$$

Aprendizagem 2022/23

Homework III – Group 107

(ist1103811, ist1103479)

2)

2)

$$f(x) = \tanh(0.5x - 2)$$

$$E = \frac{1}{2}\|z - \hat{z}\|_2^2$$

$$\eta = 0.1$$

$$x_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix},\quad x_2 = \begin{pmatrix}1\\0\\0\\-1\end{pmatrix}$$

x_1:

$$z_1^{(1)} = W^{(1)} x_1^{(0)} + b^{(1)} = \begin{pmatrix}1&1&1&1\\1&1&2&1\\1&1&1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}4\\5\\4\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}5\\6\\5\end{pmatrix}$$

$$x_1^{(1)} = f(z_1^{(1)}) = \tanh(0.5 z_1^{(1)} - 2) = \tanh\left(\begin{pmatrix}0.5\\1\\0.5\end{pmatrix}\right) = \begin{pmatrix}0.46212\\0.7616\\0.46212\end{pmatrix}$$

$$z_1^{(2)} = W^{(2)} x_1^{(1)} + b^{(2)} = \begin{pmatrix}1&4&1\\1&1&1\end{pmatrix}\begin{pmatrix}0.46212\\0.7616\\0.46212\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} =$$

$$= \begin{pmatrix}3.97064\\1.68584\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}4.97064\\2.68584\end{pmatrix}$$

$$x_1^{(2)} = f(z_1^{(2)}) = \tanh\left(\begin{pmatrix}0.48532\\-0.65708\end{pmatrix}\right) = \begin{pmatrix}0.45049\\-0.57642\end{pmatrix}$$

$$z_1^{(3)} = \begin{pmatrix}1&1\\3&1\\1&1\end{pmatrix}\begin{pmatrix}0.45049\\-0.57642\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0.87407\\1.77505\\0.87407\end{pmatrix}$$

$$x_1^{(3)} = \tanh\left(\begin{pmatrix}-1.56297\\-1.11248\\-1.56297\end{pmatrix}\right) = \begin{pmatrix}-0.9159\\-0.80494\\-0.9159\end{pmatrix}$$

x_2:

$$z_2^{(1)} = W^{(1)} x_2 + b^{(1)} = \begin{pmatrix}1&1&1&1\\1&1&2&1\\1&1&1&1\end{pmatrix}\begin{pmatrix}1\\0\\0\\-1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$x_2^{(1)} = f(z_2^{(1)}) = \tanh\left(0.5\begin{pmatrix}1\\1\\1\end{pmatrix} - 2\right) = \tanh\left(\begin{pmatrix}-1.5\\-1.5\\-1.5\end{pmatrix}\right) = \begin{pmatrix}-0.90515\\-0.90515\\-0.90515\end{pmatrix}$$

$$z_2^{(2)} = W^{(2)} x_2^{(1)} + b^{(2)} = \begin{pmatrix}1&4&1\\1&1&1\end{pmatrix}\begin{pmatrix}-0.90515\\-0.90515\\-0.90515\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} =$$

$$= \begin{pmatrix}-5.42726\\-2.71181\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-4.42726\\-1.71181\end{pmatrix}$$

$$x_2^{(2)} = \tanh\left(\begin{pmatrix}-2.21363\\-0.85591\end{pmatrix} - 2\right) = \tanh\left(\begin{pmatrix}-4.21363\\-2.85591\end{pmatrix}\right) = \begin{pmatrix}-0.99956\\-0.99341\end{pmatrix}$$

$$z_2^{(3)} = \begin{pmatrix}1&1\\3&1\\1&1\end{pmatrix}\begin{pmatrix}-0.99956\\-0.99341\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}-1.99297\\-3.99209\\-1.99297\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}-0.99297\\-2.99209\\-0.99297\end{pmatrix}$$

$$x_2^{(3)} = \tanh\left(\begin{pmatrix}-2.49649\\-3.49605\\-2.49619\end{pmatrix}\right) = \begin{pmatrix}-0.98652\\-0.99816\\-0.98652\end{pmatrix}$$

Backpropagation

$$\frac{dE}{dx^{(i)}} = \frac{d}{dx^{(i)}}\left(\frac{1}{2}\|z_t - x^{(i)}\|_2^2\right) = x^{(i)} - z_t$$

$$\frac{dx^{(i)}}{dz^{(i)}} = (\tanh(0.5z^{(i)}-2))' = 0.5(1-\tanh^2(0.5z^{(i)}-2)) = 0.5 - 0.5\tanh^2(0.5z^{(i)}-2) =$$

$$= 0.5 - (x^{(i)})^2$$

$$\frac{dz^{(i)}}{dx^{(i-1)}} = W^{(i)}$$

$$\frac{dz^{(i)}}{db^{(i)}} = 1$$

x_1:

$$\delta_1^{(3)} = \frac{dE}{dx_1^{(3)}} \circ \frac{dx_1^{(3)}}{dz_1^{(3)}} = (x_1^{(3)} - z_B) \circ (0.5 - 0.5\tanh^2(0.5z_1^{(3)}-2)) =$$

$$= (x_1^{(3)} - z_B) \circ (0.5 - 0.5(x_1^{(3)})^2) =$$

$$= \left(\begin{pmatrix}-0.9159\\-0.80494\\-0.9159\end{pmatrix} - \begin{pmatrix}-1\\1\\-1\end{pmatrix}\right) \circ \left(0.5 - 0.5\begin{pmatrix}0.83887\\0.64793\\0.83887\end{pmatrix}\right) =$$

$$= \begin{pmatrix}0.0841\\-1.80494\\0.0841\end{pmatrix} \circ \begin{pmatrix}0.08056\\0.17603\\0.08056\end{pmatrix} = \begin{pmatrix}0.00678\\-0.31772\\0.00678\end{pmatrix}$$

$$\delta_1^{(2)} = \left(\frac{dz_1^{(3)T}}{dx_1^{(2)}} \cdot \delta_1^{(3)}\right) \circ \frac{dx_1^{(2)}}{dz_1^{(2)}} = (W^{(3)T} \cdot \delta_1^{(3)}) \circ (0.5 - 0.5(x_1^{(2)})^2) =$$

$$= \left(\begin{pmatrix}1 & 3 & 1\\1 & 1 & 1\end{pmatrix} \cdot \begin{pmatrix}0.00678\\-0.31772\\0.00678\end{pmatrix}\right) \circ \left(0.5 - 0.5\begin{pmatrix}0.20294\\0.33226\end{pmatrix}\right) =$$

$$= \begin{pmatrix}-0.9396\\-0.30416\end{pmatrix} \circ \begin{pmatrix}0.39853\\0.33387\end{pmatrix} = \begin{pmatrix}-0.37446\\-0.10155\end{pmatrix}$$

$$\delta_1^{(1)} = \left(\frac{dz_1^{(2)T}}{dx_1^{(1)}} \cdot \delta_1^{(2)}\right) \circ \frac{dx_1^{(1)}}{dz_1^{(1)}} = (W^{(2)T} \cdot \delta_1^{(2)}) \circ (0.5 - 0.5(x_1^{(1)})^2) =$$

$$= \left(\begin{pmatrix}1 & 1\\4 & 1\\1 & 1\end{pmatrix}\begin{pmatrix}-0.37446\\-0.10155\end{pmatrix}\right) \circ \left(0.5 - 0.5\begin{pmatrix}0.21355\\0.58003\\0.21355\end{pmatrix}\right) = \begin{pmatrix}-0.47601\\-1.59939\\-0.47601\end{pmatrix} \circ \begin{pmatrix}0.39323\\0.20999\\0.39323\end{pmatrix} = \begin{pmatrix}-0.18718\\-0.33586\\-0.18718\end{pmatrix}$$

Homework III – Group 107

(ist1103811, ist1103479)

x_2:

$$\delta_2^{(3)} = (x_2^{(3)} - z_A) \circ (0.5 - 0.5 (x_2^{(3)})^2) =$$

$$= \left(\begin{pmatrix} -0.98652 \\ -0.99816 \\ -0.98652 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right) \circ \left(0.5 - 0.5 \begin{pmatrix} 0.97322 \\ 0.99632 \\ 0.97322 \end{pmatrix} \right) =$$

$$= \begin{pmatrix} -1.98652 \\ 0.00184 \\ 0.01348 \end{pmatrix} \circ \begin{pmatrix} 0.01339 \\ 0.00184 \\ 0.01339 \end{pmatrix} = \begin{pmatrix} -0.0266 \\ 3.39 \times 10^{-6} \\ 1.805 \times 10^{-4} \end{pmatrix}$$

$$\delta_2^{(2)} = (W^{(3)T} \cdot \delta_2^{(3)}) \circ (0.5 - 0.5 (x_2^{(2)})^2) =$$

$$= \left(\begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -0.0266 \\ 3.39 \times 10^{-6} \\ 1.805 \times 10^{-4} \end{pmatrix} \right) \circ \left(0.5 - 0.5 \begin{pmatrix} 0.99912 \\ 0.98686 \end{pmatrix} \right) =$$

$$= \begin{pmatrix} -0.02641 \\ -0.02642 \end{pmatrix} \circ \begin{pmatrix} 4.4 \times 10^{-4} \\ 6.57 \times 10^{-3} \end{pmatrix} = \begin{pmatrix} -1.16 \times 10^{-5} \\ -1.74 \times 10^{-4} \end{pmatrix}$$

$$\delta_2^{(1)} = \left(\begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1.16 \times 10^{-5} \\ -1.74 \times 10^{-4} \end{pmatrix} \right) \circ \left(0.5 - 0.5 \begin{pmatrix} 0.8193 \\ 0.8193 \\ 0.8193 \end{pmatrix} \right) =$$

$$= \begin{pmatrix} -1.86 \times 10^{-4} \\ -2.2 \times 10^{-4} \\ -1.86 \times 10^{-4} \end{pmatrix} \circ \begin{pmatrix} 0.09035 \\ 0.09035 \\ 0.09035 \end{pmatrix} = \begin{pmatrix} -1.68 \times 10^{-5} \\ -1.99 \times 10^{-5} \\ -1.68 \times 10^{-5} \end{pmatrix}$$

$$\frac{dE}{dW^{(1)}} = \delta_1^{(1)} \frac{dz_1^{(1)}}{dW^{(1)}} + \delta_2 \frac{dz_2^{(1)}}{dW^{(1)}} = \delta_1^{(1)} (x_1^{(0)T}) + \delta_2^{(1)} (x_2^{(0)T}) =$$

$$= \begin{pmatrix} -0.18718 \\ -0.33586 \\ -0.18718 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} -1.68 \times 10^{-5} \\ -1.99 \times 10^{-5} \\ -1.68 \times 10^{-5} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -1 \end{pmatrix} =$$

$$= \begin{pmatrix} -0.18718 & -0.18718 & -0.18718 & -0.18718 \\ -0.33586 & -0.33586 & -0.33586 & -0.33586 \\ -0.18718 & -0.18718 & -0.18718 & -0.18718 \end{pmatrix} + \begin{pmatrix} -1.68 \times 10^{-5} & 0 & 0 & 1.68 \times 10^{-5} \\ -1.99 \times 10^{-5} & 0 & 0 & 1.99 \times 10^{-5} \\ -1.68 \times 10^{-5} & 0 & 0 & 1.68 \times 10^{-5} \end{pmatrix} =$$

$$= \begin{pmatrix} -0.1872 & -0.18718 & -0.18718 & -0.18716 \\ -0.33588 & -0.33586 & -0.33586 & -0.33584 \\ -0.1872 & -0.18718 & -0.18718 & -0.18716 \end{pmatrix}$$

$$W^{(1)new} = W^{(1)old} - \eta \frac{dE}{dW^{(1)}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} - 0.1 \frac{dE}{dW^{(1)}} = \begin{pmatrix} 1.01872 & 1.01872 & 1.01872 & 1.01872 \\ 1.03359 & 1.03359 & 2.03359 & 1.03358 \\ 1.01872 & 1.01872 & 1.01872 & 1.01872 \end{pmatrix}$$

$$\frac{dE}{db^{(1)}} = \delta_1^{(1)} + \delta_2^{(1)} = \begin{pmatrix} -0.18718 \\ -0.33586 \\ -0.18718 \end{pmatrix} + \begin{pmatrix} -1.68 \times 10^{-5} \\ -1.99 \times 10^{-5} \\ -1.68 \times 10^{-5} \end{pmatrix} = \begin{pmatrix} -0.1872 \\ -0.33588 \\ -0.1872 \end{pmatrix}$$

$$b^{(1)new} = b^{(1)old} - \eta \frac{dE}{db^{(1)}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 0.1 \begin{pmatrix} -0.1872 \\ -0.33588 \\ -0.1872 \end{pmatrix} = \begin{pmatrix} 1.01872 \\ 1.03359 \\ 1.01872 \end{pmatrix}$$

$$\frac{dE}{dW^{(2)}} = \delta_1^{(2)}\left(x_1^{(1)T}\right) + \delta_2^{(2)}\left(x_2^{(1)T}\right) =$$

$$= \begin{pmatrix} -0.37446 \\ -0.10155 \end{pmatrix} \begin{pmatrix} 0.46212 & 0.7616 & 0.46212 \end{pmatrix} + \begin{pmatrix} -1.16\times10^{-5} \\ -1.74\times10^{-4} \end{pmatrix} \begin{pmatrix} -0.90515 & -0.90515 & -0.90515 \end{pmatrix} =$$

$$= \begin{pmatrix} -0.17305 & -0.28519 & -0.17305 \\ -0.04693 & -0.07734 & -0.04693 \end{pmatrix} + \begin{pmatrix} 1.05\times10^{-5} & 1.05\times10^{-5} & 1.05\times10^{-5} \\ 1.57\times10^{-4} & 1.57\times10^{-4} & 1.57\times10^{-4} \end{pmatrix} = \begin{pmatrix} -0.17304 & -0.28518 & -0.17304 \\ -0.04677 & -0.07718 & -0.04677 \end{pmatrix}$$

$$W^{(2)new} = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix} - 0.1 \begin{pmatrix} -0.17304 & -0.28518 & -0.17304 \\ -0.04677 & -0.07718 & -0.04677 \end{pmatrix} = \begin{pmatrix} 1.0173 & 4.02852 & 1.0173 \\ 1.00468 & 1.00772 & 1.00468 \end{pmatrix}$$

$$\frac{dE}{db^{(2)}} = \delta_1^{(2)} + \delta_2^{(2)} = \begin{pmatrix} -0.37446 \\ -0.10155 \end{pmatrix} + \begin{pmatrix} -1.16\times10^{-5} \\ -1.74\times10^{-4} \end{pmatrix} = \begin{pmatrix} -0.37447 \\ -0.10172 \end{pmatrix}$$

$$b^{(2)new} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0.1 \begin{pmatrix} -0.37447 \\ -0.10172 \end{pmatrix} = \begin{pmatrix} 1.03745 \\ 1.01017 \end{pmatrix}$$

$$\frac{dE}{dW^{(3)}} = \delta_1^{(3)}\left(x_1^{(2)T}\right) + \delta_2^{(3)}\left(x_2^{(2)T}\right) = \begin{pmatrix} 0.00678 \\ -0.31772 \\ 0.00678 \end{pmatrix} \begin{pmatrix} 0.45049 & -0.57642 \end{pmatrix} + \begin{pmatrix} -0.0266 \\ 3.39\times10^{-6} \\ 1.805\times10^{-4} \end{pmatrix} \begin{pmatrix} -0.99956 & -0.99341 \end{pmatrix} =$$

$$= \begin{pmatrix} 0.00305 & -0.00391 \\ -0.14313 & 0.18314 \\ 0.00305 & -0.00391 \end{pmatrix} + \begin{pmatrix} 0.02659 & 0.02642 \\ -3.39\times10^{-6} & -3.37\times10^{-6} \\ -1.804\times10^{-4} & -1.79\times10^{-4} \end{pmatrix} = \begin{pmatrix} 0.02964 & 0.02251 \\ -0.14313 & 0.18314 \\ 0.00287 & -0.00409 \end{pmatrix}$$

$$W^{(3)new} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{pmatrix} - 0.1 \begin{pmatrix} 0.02964 & 0.02251 \\ -0.14313 & 0.18314 \\ 0.00287 & -0.00409 \end{pmatrix} = \begin{pmatrix} 0.99704 & 0.99775 \\ 3.01431 & 0.98169 \\ 0.99971 & 1.00041 \end{pmatrix}$$

$$\frac{dE}{db^{(3)}} = \delta_1^{(3)} + \delta_2^{(3)} = \begin{pmatrix} 0.00678 \\ -0.31772 \\ 0.00678 \end{pmatrix} + \begin{pmatrix} -0.0266 \\ 3.39\times10^{-6} \\ 1.805\times10^{-4} \end{pmatrix} = \begin{pmatrix} -0.01982 \\ -0.31772 \\ 0.00696 \end{pmatrix}$$

$$b^{(3)new} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 0.1 \begin{pmatrix} -0.01982 \\ -0.31772 \\ 0.00696 \end{pmatrix} = \begin{pmatrix} 1.00198 \\ 1.031772 \\ 0.9993 \end{pmatrix}$$

Aprendizagem 2022/23

Homework III – Group 107

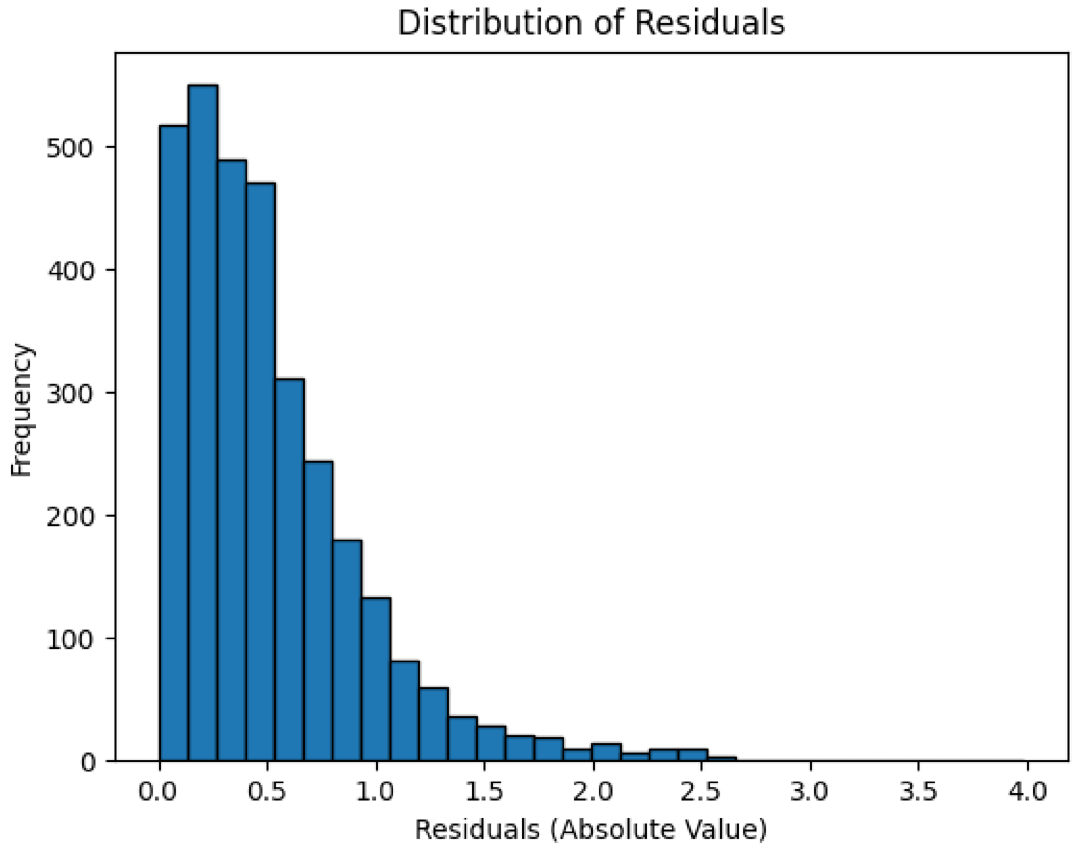
(ist1103811, ist1103479)

II. Programming and critical analysis

1)a)

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_absolute_error
import matplotlib.pyplot as plt

# Load the dataset
wine_data = pd.read_csv("winequality-red.csv", sep=";")

# Define features and target variable
X = wine_data.drop("quality", axis=1)
y = wine_data["quality"]

# Initialize variables to store residuals from 10 runs
residuals = []

# Set the random seeds for reproducibility
random_seeds = range(1, 11)

# Split the dataset into training and test sets (80-20 split)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ran-
dom_state=0)

# Perform 10 runs
for random_seed in random_seeds:

    # Create and train the MLP regressor
    mlp_regressor = MLPRegressor(hidden_layer_sizes=(10, 10), activation='relu',
                                 random_state=random_seed, early_stopping=True,
                                 validation_fraction=0.2)

    mlp_regressor.fit(X_train, y_train)

    # Predict on the test set
    y_pred = mlp_regressor.predict(X_test)

    # Calculate residuals
    residual = np.abs(y_test - y_pred)
    residuals.extend(residual)

# Plot the distribution of residuals using a histogram
plt.hist(residuals, bins=30, edgecolor='k')
plt.xlabel('Residuals (Absolute Value)')
```

```
plt.ylabel('Frequency')
plt.title('Distribution of Residuals')
plt.savefig("Ex1IMAGE")
```



2)

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_absolute_error
import matplotlib.pyplot as plt

# Load the dataset
wine_data = pd.read_csv("winequality-red.csv", sep=";")

# Define features and target variable
X = wine_data.drop("quality", axis=1)
y = wine_data["quality"]

# Initialize variables to store residuals from 10 runs
residuals = []

# Set the random seeds for reproducibility
random_seeds = range(1, 11)

# Initialize variables to store MAE before and after rounding and bounding
mae_before = []
mae_after = []

# Define rounding and bounding function
def round_and_bound(predictions):
    rounded_predictions = np.round(predictions)
    bounded_predictions = np.clip(rounded_predictions, a_min=3, a_max=8)  # Ad-
just bounds as needed
    return bounded_predictions

# Perform 10 runs
for random_seed in random_seeds:
    # Split the dataset into training and test sets (80-20 split)
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ran-
dom_state=0)

    # Create and train the MLP regressor
    mlp_regressor = MLPRegressor(hidden_layer_sizes=(10, 10), activation='relu',
                                 random_state=random_seed, early_stopping=True,
                                 validation_fraction=0.2)
    mlp_regressor.fit(X_train, y_train)
```

```python
    # Predict on the test set
    y_pred = mlp_regressor.predict(X_test)

    # Calculate residuals before rounding and bounding
    residual_before = np.abs(y_test - y_pred)
    mae_before.append(mean_absolute_error(y_test, y_pred))

    # Round and bound estimates
    y_pred_rounded_and_bounded = round_and_bound(y_pred)

    # Calculate residuals after rounding and bounding
    residual_after = np.abs(y_test - y_pred_rounded_and_bounded)
    mae_after.append(mean_absolute_error(y_test, y_pred_rounded_and_bounded))

    # Store residuals
    residuals.extend(residual_after)


# Calculate and print the average MAE before and after rounding and bounding
avg_mae_before = np.mean(mae_before)
avg_mae_after = np.mean(mae_after)
print(f"Average MAE before rounding and bounding: {avg_mae_before}")
print(f"Average MAE after rounding and bounding: {avg_mae_after}")
```

Average MAE before rounding and bounding: 0.5097171955009514

Average MAE after rounding and bounding: 0.43875000000000003

3)

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
from math import sqrt

# Load the dataset
wine_data = pd.read_csv("winequality-red.csv", sep=";")

# Define features and target variable
X = wine_data.drop("quality", axis=1)
y = wine_data["quality"]

# Initialize variables to store RMSE for different numbers of iterations and for
early stopping
rmse_values = []

# Set the random seeds for reproducibility
random_seeds = range(1, 11)

# Define different numbers of iterations and early stopping as 1 to include it
iterations = [1, 20, 50, 100, 200]

# Perform 10 runs for each number of iterations
for num_iterations in iterations:
    rmse_per_iteration = []  # To store RMSE for each run with the given number
of iterations

    for random_seed in random_seeds:
        # Split the dataset into training and test sets (80-20 split)
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=0)

        if num_iterations == 1:
            # Create and train the MLP regressor with early stopping
            mlp_regressor = MLPRegressor(hidden_layer_sizes=(10, 10),
                                         activation='relu',
                                         random_state=random_seed,
                                         early_stopping=True,
                                         validation_fraction=0.2)
```

```
        else:
            # Create and train the MLP regressor with the specified number of it-
erations
            mlp_regressor = MLPRegressor(hidden_layer_sizes=(10, 10),
                                         activation='relu',
                                         random_state=random_seed,
                                         max_iter=num_iterations,
                                         solver='adam')

        mlp_regressor.fit(X_train, y_train)

        # Predict on the test set
        y_pred = mlp_regressor.predict(X_test)

        # Calculate RMSE
        rmse = sqrt(mean_squared_error(y_test, y_pred))
        rmse_per_iteration.append(rmse)

    # Calculate the average RMSE for the given number of iterations or early
stopping
    avg_rmse = np.mean(rmse_per_iteration)
    rmse_values.append(avg_rmse)


# Print RMSE values for each number of iterations or early stopping
for num_iterations, rmse in zip(iterations, rmse_values):
    if num_iterations == 1:
        print(f"Early Stopping, Average RMSE: {rmse}")
    else:
        print(f"Number of Iterations: {num_iterations}, Average RMSE: {rmse}")
```

Early Stopping, Average RMSE: 0.6706527958221328

Number of Iterations: 20, Average RMSE: 1.4039789509925442

Number of Iterations: 50, Average RMSE: 0.7996073631460568

Number of Iterations: 100, Average RMSE: 0.6940361469112143

Number of Iterations: 200, Average RMSE: 0.6554543932216474

4)

Com o intuito de tentar melhorar os nossos resultados, podemos utilizar o "early stopping". Este método não se baseia num número específico de iterações mas sim no momento em que a performance do modelo começa a baixar. Existem algumas vantagens ao usar early stoppping: reduz a ocorrência de overfitting e pode até economizar algum tempo de treino, se a performance começar a reduzir cedo. No entanto, também existem desvantagens, principalmente o risco de haver um "stop" prematuro, causando um problema de underfitting. Assim, ao usarmos early stopping, estaríamos a prevenir o overfitting, porém teríamos de ter em conta os casos de "stops" prematuros que iriam causar underfitting (embora que raros) e performances subóptimas.

Ao analisar os valores do exercício anterior, em que o RMSE é analisado através de um número fixos de iterações, percebemos que este vai diminuindo com o aumento de iterações. Um ponto importante a referir é que ao calcular o average RMSE com early stopping (dando um valor de 0.6706527958221328) podemos compara-lo com os RMSE de iterações fixas. Vemos que o RMSE do early stopping é menor que todos os outros execto em 200 iterações. Isto quer dizer que nas iteraçoes de 20,50 e 100 estamos presentes um problema de underfitting, o modelo não teve tempo para convergir. Outra observação que também podemos fazer é em relação aos testing e validation sets provenientes do early stopping. Para além das desvantagens anteriormente apresentadas, o early stopping é sensível à escolha de testing e validation sets. Se o validition set não for suficientemente representativo do testing set o modelo pode sofrer de underfitting ou overfitting. Assim, e assumindo que o número máximo de iterações no early stopping é 200 podemos comparar a performance deste método com o de 200 iterações fixas. Visto que o de 200 iterações fixas é menor podemos concluir que o validation set usado no early stopping é subóptimo, havendo um mais representativo do testing set.

END