Homework III – Group 075

(ist1103860, ist1103883)

I. Pen-and-paper

1) $\left\{ \begin{pmatrix} 0.7 \\ -0.3 \end{pmatrix}, \begin{pmatrix} 0.4 \\ 0.5 \end{pmatrix}, \begin{pmatrix} -0.2 \\ 0.8 \end{pmatrix}, \begin{pmatrix} -0.4 \\ 0.3 \end{pmatrix} \right\}$ $\Longrightarrow$

Targets: $(0.8, 0.6, 0.3, 0.3)$

	y_1	y_2	output (z)
x_1	0.7	-0.3	0.8
x_2	0.4	0.5	0.6
x_3	-0.2	0.8	0.3
x_4	-0.4	0.3	0.3

(a) $\phi_j(x) = \exp\left(-\dfrac{\|x - c_j\|^2}{2}\right)$, $\left\{ c_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, c_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, c_3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$, $\lambda = 0.1$

	$i=1$	$i=2$	$i=3$	$i=4$
$\phi_1(x_i)$	$\exp\left(-\frac{\|x_1 - c_1\|^2}{2}\right) = \exp\left(-\frac{0.7^2 + (-0.3)^2}{2}\right) \simeq$ $\simeq 0.74826$	$\exp\left(-\frac{\|x_2 - c_1\|^2}{2}\right) = \exp\left(-\frac{0.4^2 + 0.5^2}{2}\right) \simeq$ $\simeq 0.81465$	$\exp\left(-\frac{\|x_3 - c_1\|^2}{2}\right) = \exp\left(-\frac{(-0.2)^2 + 0.8^2}{2}\right) \simeq$ $\simeq 0.71177$	$\exp\left(-\frac{\|x_4 - c_1\|^2}{2}\right) = \exp\left(-\frac{(-0.4)^2 + 0.3^2}{2}\right) =$ $\simeq 0.88250$
$\phi_2(x_i)$	$\exp\left(-\frac{\|x_1 - c_2\|^2}{2}\right) = \exp\left(-\frac{(0.7-1)^2 + (-0.3+1)^2}{2}\right) \simeq$ $\simeq \exp\left(-\frac{(-0.3)^2 + 0.7^2}{2}\right) \simeq$ $\simeq 0.74826$	$\exp\left(-\frac{\|x_2 - c_2\|^2}{2}\right) = \exp\left(-\frac{(0.4-1)^2 + (0.5+1)^2}{2}\right)$ $= \exp\left(-\frac{(-0.6)^2 + 1.5^2}{2}\right) \simeq$ $\simeq 0.27117$	$\exp\left(-\frac{\|x_3 - c_2\|^2}{2}\right) = \exp\left(-\frac{(-0.2-1)^2 + (0.8+1)^2}{2}\right)$ $= \exp\left(-\frac{(-1.2)^2 + 1.8^2}{2}\right)$ $\simeq 0.09633$	$\exp\left(-\frac{\|x_4 - c_2\|^2}{2}\right) = \exp\left(-\frac{(-0.4-1)^2 + (0.3+1)^2}{2}\right)$ $= \exp\left(-\frac{(-1.4)^2 + 1.3^2}{2}\right)$ $\simeq 0.16122$
$\phi_3(x_i)$	$\exp\left(-\frac{\|x_1 - c_3\|^2}{2}\right) = \exp\left(-\frac{(0.7+1)^2 + (-0.3-1)^2}{2}\right)$ $= \exp\left(-\frac{1.7^2 + (-1.3)^2}{2}\right)$ $\simeq 0.10127$	$\exp\left(-\frac{\|x_2 - c_3\|^2}{2}\right) = \exp\left(-\frac{(0.4+1)^2 + (0.5-1)^2}{2}\right)$ $= \exp\left(-\frac{1.4^2 + (-0.5)^2}{2}\right)$ $\simeq 0.33121$	$\exp\left(-\frac{\|x_3 - c_3\|^2}{2}\right) = \exp\left(-\frac{(-0.2+1)^2 + (0.8-1)^2}{2}\right)$ $= \exp\left(-\frac{0.8^2 + (-0.2)^2}{2}\right)$ $\simeq 0.71177$	$\exp\left(-\frac{\|x_4 - c_3\|^2}{2}\right) = \exp\left(-\frac{(-0.4+1)^2 + (0.3-1)^2}{2}\right)$ $= \exp\left(-\frac{0.6^2 + (-0.7)^2}{2}\right)$ $\simeq 0.65377$

Ridge: $w = (X^T X + \lambda I)^{-1} X^T z$, tendo em conta $\phi_j(x)$ tem-se que

$$w = (\phi^T \phi + \lambda I)^{-1} \phi^T z$$

bias

- $\phi = \begin{pmatrix} 1 & 0.74826 & 0.74826 & 0.10127 \\ 1 & 0.81465 & 0.27117 & 0.33121 \\ 1 & 0.71177 & 0.09633 & 0.71177 \\ 1 & 0.88250 & 0.16122 & 0.65377 \end{pmatrix}$ • $z = \begin{pmatrix} 0.8 \\ 0.6 \\ 0.3 \\ 0.3 \end{pmatrix}$

observações com as transformações aplicadas

- $$\phi^T\phi = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0.74826 & 0.81465 & 0.71177 & 0.88250 \\ 0.74826 & 0.27117 & 0.09633 & 0.16122 \\ 0.10127 & 0.33121 & 0.71177 & 0.65377 \end{pmatrix} \begin{pmatrix} 1 & 0.74826 & 0.74826 & 0.10127 \\ 1 & 0.81465 & 0.27117 & 0.33121 \\ 1 & 0.71177 & 0.09633 & 0.71177 \\ 1 & 0.88250 & 0.16122 & 0.65377 \end{pmatrix}$$

$$= \begin{pmatrix} 4 & 3.15718 & 1.27698 & 1.79802 \\ 3.15718 & 2.50897 & 0.99165 & 1.42916 \\ 1.27698 & 0.99165 & 0.66870 & 0.33955 \\ 1.79802 & 1.42916 & 0.33955 & 1.05399 \end{pmatrix}$$

- $$\phi^T\phi + \lambda I = \phi^T\phi + 0.1 I = \begin{pmatrix} 4 & 3.15718 & 1.27698 & 1.79802 \\ 3.15718 & 2.50897 & 0.99165 & 1.42916 \\ 1.27698 & 0.99165 & 0.66870 & 0.33955 \\ 1.79802 & 1.42916 & 0.33955 & 1.05399 \end{pmatrix} +$$

$$+ \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix} = \begin{pmatrix} 4.1 & 3.15718 & 1.27698 & 1.79802 \\ 3.15718 & 2.60897 & 0.99165 & 1.42916 \\ 1.27698 & 0.99165 & 0.76870 & 0.33955 \\ 1.79802 & 1.42916 & 0.33955 & 1.15399 \end{pmatrix}$$

- $$(\phi^T\phi + \lambda I)^{-1} = (\phi^T\phi + 0.1 I)^{-1} = \begin{pmatrix} 4.54826 & -3.77682 & -1.86117 & -1.86155 \\ -3.77682 & 5.98285 & -0.88543 & -1.26432 \\ -1.86117 & -0.88543 & 4.33276 & 2.72156 \\ -1.86155 & -1.26432 & 2.72156 & 4.53204 \end{pmatrix}$$

- $$(\phi^T\phi + \lambda I)^{-1}\phi^T = \begin{pmatrix} 4.54826 & -3.77682 & -1.86117 & -1.86155 \\ -3.77682 & 5.98285 & -0.88543 & -1.26432 \\ -1.86117 & -0.88543 & 4.33276 & 2.72156 \\ -1.86155 & -1.26432 & 2.72156 & 4.53204 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0.74826 & 0.81465 & 0.71177 & 0.88250 \\ 0.74826 & 0.27117 & 0.09633 & 0.16122 \\ 0.10127 & 0.33121 & 0.71177 & 0.65377 \end{pmatrix} = \begin{pmatrix} 0.14105 & 0.35022 & 0.35575 & -0.30185 \\ -0.09064 & 0.43823 & -0.50361 & 0.53370 \\ 0.99394 & -0.50615 & -0.13690 & -0.16477 \\ -0.31222 & -0.65246 & 0.72647 & 0.42436 \end{pmatrix}$$

- $$(\phi^T\phi + \lambda I)^{-1}\phi^T z = \begin{pmatrix} 0.14105 & 0.35022 & 0.35575 & -0.30185 \\ -0.09064 & 0.43823 & -0.50361 & 0.53370 \\ 0.99394 & -0.50615 & -0.13690 & -0.16477 \\ -0.31222 & -0.65246 & 0.72647 & 0.42436 \end{pmatrix} \begin{pmatrix} 0.8 \\ 0.6 \\ 0.3 \\ 0.3 \end{pmatrix} =$$

$$= \begin{bmatrix} 0.33914 & 0.19945 & 0.40096 & -0.296 \end{bmatrix}^T = w$$

Tendo calculado o vetor w, temos que os valores dos pesos para a regressão são:

$w_0 = 0.33914$ $\quad w_1 = 0.19945$ $\quad w_2 = 0.40096$ $\quad w_3 = -0.29600$

Sendo a regressão dada por: $\hat{z}(x, w) = w_0 + w_1 x + w_2 x^2 + w_3 x^3$,

temos: $\hat{z}(x) = 0.33914 + 0.19945x + 0.40096x^2 - 0.29600x^3$

(b) Root Mean Square Error (RMSE): $RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(z_i - \hat{z}_i)^2}$

$m = 4$, visto que temos 4 observações

Usando $\hat{z}(x) = 0.33914 + 0.19945x + 0.40096x^2 - 0.29600x^3$, podemos calcular $\hat{z}$ para as 4 observações fazendo a seguinte multiplicação de matrizes:

- $z = \begin{pmatrix} 0.8 \\ 0.6 \\ 0.3 \\ 0.3 \end{pmatrix}$

- $\hat{z} = \phi w = \begin{pmatrix} 1 & 0.74826 & 0.74826 & 0.10127 \\ 1 & 0.81465 & 0.27117 & 0.33121 \\ 1 & 0.71177 & 0.09633 & 0.71177 \\ 1 & 0.88250 & 0.16122 & 0.65377 \end{pmatrix} \times \begin{pmatrix} 0.33914 \\ 0.19945 \\ 0.40096 \\ 0.29600 \end{pmatrix} = \begin{pmatrix} 0.75844 \\ 0.51232 \\ 0.30905 \\ 0.38629 \end{pmatrix}$

- $\sum_{i=1}^{4}(z_i - \hat{z}_i)^2 = (0.8 - 0.75844)^2 + (0.6 - 0.51232)^2 + (0.3 - 0.30905)^2 + (0.3 - 0.38629)^2$
  $\simeq 0.01694$

$RMSE = \sqrt{\frac{1}{4}\sum_{i=1}^{4}(z_i - \hat{z}_i)^2} = \frac{1}{2}\sqrt{0.01694} \simeq 0.06508$

R: RMSE = 0.06508

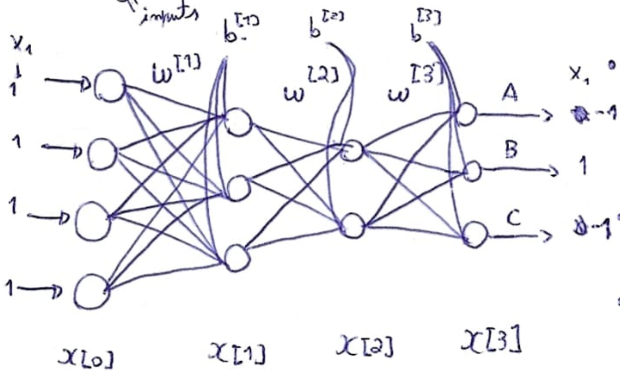
2) outcomes: A, B e C

$W^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$, $b^{[1]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $W^{[2]} = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}$, $b^{[2]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $W^{[3]} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}$, $b^{[3]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

$W^{[1]} \downarrow$ ~~3x3~~ 3×4 ← inputs

$W^{[2]} \downarrow$ 2×3

$W^{[3]} \downarrow$ 3×2

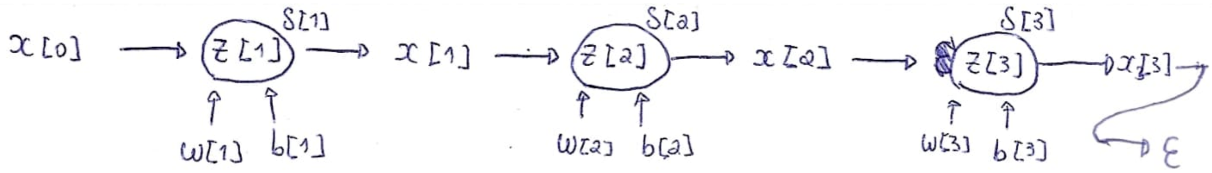


- função de ativação: $f(x) = \frac{e^{0.5x-2} - e^{-0.5x+2}}{e^{0.5x-2} + e^{-0.5x+2}} = \tanh(0.5x - 2)$

- $\eta = 0.1$

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

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



1) Propagation:

→ Observação $x_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$

$x^{[0]} = [1 \quad 1 \quad 1 \quad 1]^T$

• $z[1] = w^{[1]} x^{[0]} + b^{[1]} = \begin{pmatrix}1&1&1&1\\1&1&2&1\\1&1&1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} =$

$= \begin{pmatrix}4\\5\\4\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}5\\6\\5\end{pmatrix}$

• $x[1] = f(z^{[1]}) = f\left(\begin{bmatrix}5\\6\\5\end{bmatrix}\right) = \begin{pmatrix}0.46212\\0.76159\\0.46212\end{pmatrix}$

• $z^{[2]} = w^{[2]} x^{[1]} + b^{[2]} = \begin{pmatrix}1&4&1\\1&1&1\end{pmatrix}\begin{pmatrix}0.46212\\0.76159\\0.46212\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} =$

$= \begin{pmatrix}3.97061\\1.68583\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}4.97061\\2.68583\end{pmatrix}$

• $x^{[2]} = f(z^{[2]}) = f\left(\begin{bmatrix}4.97061\\2.68583\end{bmatrix}\right) = \begin{pmatrix}0.45048\\-0.57642\end{pmatrix}$

• $z^{[3]} = w^{[3]} x^{[2]} + b^{[3]} = \begin{pmatrix}1&1\\3&1\\1&1\end{pmatrix}\begin{pmatrix}0.45048\\-0.57642\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} =$

$= \begin{pmatrix}-0.12594\\0.77503\\-0.12594\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0.87406\\1.77503\\0.87406\end{pmatrix}$

• $x^{[3]} = f(z^{[3]}) = f\left(\begin{bmatrix}0.87406\\1.77503\\0.87406\end{bmatrix}\right) = \begin{pmatrix}-0.91590\\-0.80494\\-0.91590\end{pmatrix}$

→ Observação $X_2 = \begin{pmatrix}1\\0\\0\\-1\end{pmatrix}$

$x^{[0]} = [1 \quad 0 \quad 0 \quad -1]^T$

• $z^{[1]} = w^{[1]} x^{[0]} + b^{[1]} = \begin{pmatrix}1&1&1&1\\1&1&2&1\\1&1&1&1\end{pmatrix}\begin{pmatrix}1\\0\\0\\-1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$

- $x^{[1]} = f(z^{[1]}) = f\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{pmatrix}-0.90515\\-0.90515\\-0.90515\end{pmatrix}$

- $z^{[2]} = W^{[2]}x^{[1]} + b^{[2]} = \begin{pmatrix}1 & 4 & 1\\1 & 1 & 1\end{pmatrix}\begin{pmatrix}-0.90515\\-0.90515\\-0.90515\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-5.43089\\-2.71544\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} =$
$= \begin{pmatrix}-4.43089\\-1.71544\end{pmatrix}$

- $x^{[2]} = f(z^{[2]}) = f\left(\begin{bmatrix}-4.43089\\-1.71544\end{bmatrix}\right) = \begin{pmatrix}-0.99956\\-0.99343\end{pmatrix}$

- $z^{[3]} = W^{[3]}x^{[2]} + b^{[3]} = \begin{pmatrix}1 & 1\\3 & 1\\1 & 1\end{pmatrix}\begin{pmatrix}-0.99956\\-0.99343\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}-1.99300\\-3.99212\\-1.99300\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} =$
$= \begin{pmatrix}-0.99300\\-2.99212\\-0.99300\end{pmatrix}$

- $x^{[3]} = f(z^{[3]}) = f\left(\begin{bmatrix}-0.99300\\-2.99212\\-0.99300\end{bmatrix}\right) = \begin{pmatrix}-0.98652\\-0.99816\\-0.98652\end{pmatrix}$

② BackPropagation

- $E(w) = \frac{1}{2}\|z - \hat{z}\|_2^2 = \frac{1}{2}\|z - x^{[3]}\|_2^2$, como $\|x\|_p = \left(\sum_{i=1}^{m} x_i^p\right)^{1/p}$ tem-se:

$$E(w) = \frac{1}{2}\left(\left(\sum_{i=1}^{m}(z_i - x_i^{[3]})^2\right)^{1/2}\right)^2 = \frac{1}{2}\sum_{i=1}^{m}(z_i - x_i^{[3]})^2 \xrightarrow{\frac{\partial}{\partial x^{[3]}}} \frac{1}{2}\sum_{i=1}^{m} 2(z_i - x_i^{[3]})\frac{\partial}{\partial x^{[3]}}(z_i - x_i^{[3]}) =$$
$$= -\sum_{i=1}^{m}(z_i - x_i^{[3]})$$

Para uma dada observação:

$$\frac{\partial E}{\partial x^{[3]}} = -(z - x^{[3]}) = x^{[3]} - z$$

Derivadas

- $\frac{\partial E}{\partial x^{[3]}} = x^{[3]} - z$
- $\frac{\partial x^{[i]}}{\partial z^{[i]}} = 0.5 - 0.5 f^2(z^{[i]})$
- $\frac{\partial z^{[i]}}{\partial W^{[i]}} = x^{[i-1]}$
- $\frac{\partial z^{[i]}}{\partial x^{[i-1]}} = W^{[i]}$
- $\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$

C. Aux

- $(\tanh(x))' = 1 - \tanh^2(x)$
- $(\tanh(0.5x - 2))' = (1 - \tanh^2(0.5x - 2)) \times 0.5 = 0.5 - 0.5\tanh^2(0.5x - 2) =$
$= 0.5 - 0.5 f^2(x)$

Deltas $\rightarrow$ Observação $X_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, atendendo que $f: \mathbb{R} \rightarrow [-1,1]$ e que X_1 é classificada como B, tem-se que $z = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$

$$\bullet\ \delta^{[3]} = \frac{\partial E}{\partial x^{[3]}} \circ \frac{\partial x^{[3]}}{\partial z^{[3]}} = (x^{[3]} - z) \circ (0.5 - 0.5 f^2(z^{[3]})) = \left(\begin{pmatrix} -0.91590 \\ -0.80494 \\ -0.91590 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} \right) \circ \Bigg(0.5 - 0.5 \underbrace{f^2\left(\begin{pmatrix} 0.37406 \\ 1.77503 \\ 0.37406 \end{pmatrix} \right)}_{(x^{[3]})^2}\Bigg) = \begin{pmatrix} 0.0841 \\ -1.80494 \\ 0.0841 \end{pmatrix} \circ \begin{pmatrix} 0.5 - 0.5 \times (-0.91590)^2 \\ 0.5 - 0.5 \times (-0.80494)^2 \\ 0.5 - 0.5 \times (-0.91590)^2 \end{pmatrix} = \begin{pmatrix} 0.0841 \\ -1.80494 \\ 0.0841 \end{pmatrix} \circ \begin{pmatrix} 0.08056 \\ 0.17604 \\ 0.08056 \end{pmatrix} = \begin{pmatrix} 0.00678 \\ -0.31773 \\ 0.00678 \end{pmatrix}$$

$$\bullet\ \delta^{[2]} = \left(\frac{\partial z^{[3]}}{\partial x^{[2]}}^T \cdot \delta^{[3]} \right) \circ \frac{\partial x^{[2]}}{\partial z^{[2]}} = \left(w^{[3]T} \cdot \delta^{[3]} \right) \circ \Big(0.5 - 0.5 \underbrace{f^2(z^{[2]})}_{(x^{[2]})^2}\Big) =$$

$$= \left(\begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.00678 \\ -0.31773 \\ 0.00678 \end{pmatrix} \right) \circ \begin{pmatrix} 0.5 - 0.5 \times 0.45048^2 \\ 0.5 - 0.5 \times (-0.57642)^2 \\ \cancel{0.5 - 0.5 \times} \end{pmatrix} = \begin{pmatrix} -0.93965 \\ -0.30418 \end{pmatrix} \circ \begin{pmatrix} 0.39853 \\ 0.333870 \end{pmatrix} = \begin{pmatrix} -0.37448 \\ -0.10156 \end{pmatrix}$$

$$\bullet\ \delta^{[1]} = \left(\frac{\partial z^{[2]}}{\partial x^{[1]}}^T \cdot \delta^{[2]} \right) \circ \frac{\partial x^{[1]}}{\partial z^{[1]}} = \left(w^{[2]T} \cdot \delta^{[2]} \right) \circ \Big(0.5 - 0.5 \underbrace{f^2(z^{[1]})}_{(x^{[1]})^2}\Big) =$$

$$= \left(\begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -0.37448 \\ -0.10156 \end{pmatrix} \right) \circ \begin{pmatrix} 0.5 - 0.5 \times 0.46212^2 \\ 0.5 - 0.5 \times 0.76159^2 \\ 0.5 - 0.5 \times 0.46212^2 \end{pmatrix} = \begin{pmatrix} -0.47604 \\ -1.59949 \\ -0.47604 \end{pmatrix} \circ \begin{pmatrix} 0.39322 \\ 0.20999 \\ 0.39322 \end{pmatrix} = \begin{pmatrix} -0.18719 \\ -0.33587 \\ -0.18719 \end{pmatrix}$$

$\rightarrow$ Calculando agora $\delta^{[1]}$, $\delta^{[2]}$ e $\delta^{[3]}$ para $x_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$, sendo $f: \mathbb{R} \rightarrow [-1,1]$ e X_2 classificada como A, tem-se que $z = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$

$$\bullet\ \delta^{[3]} = \frac{\partial E}{\partial x^{[3]}} \circ \frac{\partial x^{[3]}}{\partial z^{[3]}} = (x^{[3]} - z) \circ (0.5 - 0.5 f^2(z^{[3]})) = \left(\begin{pmatrix} -0.98652 \\ -0.99816 \\ -0.98652 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right) \circ \Bigg(0.5 - 0.5 \underbrace{f^2\left(\begin{pmatrix} -0.99300 \\ -2.99212 \\ -0.99300 \end{pmatrix} \right)}_{(x^{[3]})^2}\Bigg) = \begin{pmatrix} -1.98652 \\ 1.84000 \times 10^{-3} \\ 1.34800 \times 10^{-2} \end{pmatrix} \circ \begin{pmatrix} 0.5 - 0.5 \times (-0.98652)^2 \\ 0.5 - 0.5 \times (-0.99816)^2 \\ 0.5 - 0.5 \times (-0.98652)^2 \end{pmatrix} = \begin{pmatrix} -1.98652 \\ 1.84000 \times 10^{-3} \\ 1.34800 \times 10^{-2} \end{pmatrix} \circ \begin{pmatrix} 0.01339 \\ 0.00183 \\ 0.01339 \end{pmatrix} = \begin{pmatrix} -0.0266 \\ 3.370 \times 10^{-6} \\ 0.00018 \end{pmatrix}$$

$$\bullet\ \delta^{[2]} = \left(\frac{\partial z^{[3]}}{\partial x^{[2]}}^T \cdot \delta^{[3]} \right) \circ \frac{\partial x^{[2]}}{\partial z^{[2]}} = \left(w^{[3]T} \cdot \delta^{[3]} \right) \circ \Big(0.5 - 0.5 \underbrace{f^2(z^{[2]})}_{(x^{[2]})^2}\Big) =$$

$$= \left(\begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -0.0266 \\ 3.370 \times 10^{-6} \\ 0.00018 \end{pmatrix} \right) \circ \begin{pmatrix} 0.5 - 0.5 \times (-0.99956)^2 \\ 0.5 - 0.5 \times (-0.99343)^2 \end{pmatrix} = \begin{pmatrix} -0.02641 \\ -0.02641 \end{pmatrix} \circ \begin{pmatrix} 0.00044 \\ 0.00655 \end{pmatrix} = \begin{pmatrix} -1.1509 \times 10^{-5} \\ -1.72899 \times 10^{-4} \end{pmatrix}$$

$$\bullet\ \delta^{[1]} = \left(\frac{\partial z^{[2]}}{\partial x^{[1]}}^T \cdot \delta^{[2]} \right) \circ \left(\frac{\partial x^{[1]}}{\partial z^{[1]}} \right) = \left(w^{[2]T} \cdot \delta^{[2]} \right) \circ \Big(0.5 - 0.5 \underbrace{f^2(z^{[1]})}_{(x^{[1]})^2}\Big) =$$

$$= \left(\begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1.1509 \times 10^{-5} \\ -1.72899 \times 10^{-4} \end{pmatrix} \right) \circ \begin{pmatrix} 0.5 - 0.5 \times (-0.90515)^2 \\ 0.5 - 0.5 \times (-0.90515)^2 \\ 0.5 - 0.5 \times (-0.90515)^2 \end{pmatrix} = \begin{pmatrix} -0.00018 \\ -0.00022 \\ -0.00018 \end{pmatrix} \circ \begin{pmatrix} 0.09035 \\ 0.09035 \\ 0.09035 \end{pmatrix} = \begin{pmatrix} -1.66619 \times 10^{-5} \\ -1.97816 \times 10^{-5} \\ -1.66619 \times 10^{-5} \end{pmatrix}$$

- $\Delta w^{[i]} = -\eta \dfrac{\partial E}{\partial w^{[i]}}$, $i \in \{1,2\}$ • $\Delta b^{[i]} = -\eta \dfrac{\partial E}{\partial b^{[i]}}$, $i \in \{1,2\}$

- $\Delta w^{[i]} = -0.1 \left(\delta^{[i]} \cdot \dfrac{\partial x^{[i]}}{\partial w^{[i]}}^T \right) = -0.1 \left(\delta^{[i]} \cdot x^{[i-1]T} \right)$

- $\Delta b^{[i]} = -0.1 \left(\delta^{[i]} \cdot \underbrace{\dfrac{\partial z^{[i]}}{\partial b^{[i]}}}_{1}^{T} \right) = -0.1\ \delta^{[i]}$

Continuando para a observação X_1:

- $\Delta w^{[1]} = -0.1 \left(\delta^{[1]} \cdot (x^{[0]})^T \right) = -0.1 \left(\begin{pmatrix} -0.18719 \\ -0.33587 \\ -0.18719 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \right) = -0.1 \begin{pmatrix} -0.18719 & -0.18719 & -0.18719 & -0.18719 \\ -0.33587 & -0.33587 & -0.33587 & -0.33587 \\ -0.18719 & -0.18719 & -0.18719 & -0.18719 \end{pmatrix} = \begin{pmatrix} 0.01872 & 0.01872 & 0.01872 & 0.01872 \\ 0.03359 & 0.03359 & 0.03359 & 0.03359 \\ 0.01872 & 0.01872 & 0.01872 & 0.01872 \end{pmatrix}$

- $\Delta b^{[1]} = -0.1\ \delta^{[1]} = -0.1 \begin{pmatrix} -0.18719 \\ -0.33587 \\ -0.18719 \end{pmatrix} = \begin{pmatrix} 0.01872 \\ 0.03359 \\ 0.01872 \end{pmatrix}$

- $\Delta w^{[2]} = -0.1 \left(\delta^{[2]} \cdot (x^{[1]})^T \right) = -0.1 \left(\begin{pmatrix} -0.37448 \\ -0.10156 \end{pmatrix} \begin{pmatrix} 0.46212 & 0.76159 & 0.46212 \end{pmatrix} \right) = -0.1 \begin{pmatrix} -0.17305 & -0.28520 & -0.17305 \\ -0.04693 & -0.07735 & -0.04693 \end{pmatrix} = \begin{pmatrix} 0.01731 & 0.02852 & 0.01731 \\ 0.00469 & 0.00774 & 0.00469 \end{pmatrix}$

- $\Delta b^{[2]} = -0.1\ \delta^{[2]} = -0.1 \begin{pmatrix} -0.37448 \\ -0.10156 \end{pmatrix} = \begin{pmatrix} 0.03745 \\ 0.01016 \end{pmatrix}$

- $\Delta w^{[3]} = -0.1 \left(\delta^{[3]} \cdot (x^{[2]})^T \right) = -0.1 \left(\begin{pmatrix} 0.00678 \\ -0.31773 \\ 0.00678 \end{pmatrix} \begin{pmatrix} 0.45048 & -0.57642 \end{pmatrix} \right) = -0.1 \begin{pmatrix} 0.00305 & -0.00391 \\ -0.14313 & 0.18315 \\ 0.00305 & -0.00391 \end{pmatrix} = \begin{pmatrix} -0.00031 & 0.00039 \\ 0.01431 & -0.01832 \\ -0.00031 & 0.00039 \end{pmatrix}$

- $\Delta b^{[3]} = -0.1\ \delta^{[3]} = -0.1 \begin{pmatrix} 0.00678 \\ -0.31773 \\ 0.00678 \end{pmatrix} = \begin{pmatrix} -0.00068 \\ 0.03177 \\ -0.00068 \end{pmatrix}$

Agora, repetindo para a observação X_2:

$$\bullet\ \Delta W^{[1]} = -0.1\left(\delta^{[1]} \cdot (x^{[0]})^T\right) = -0.1\left(\begin{pmatrix} -1.66619\times10^{-5} \\ -1.97816\times10^{-5} \\ -1.66619\times10^{-5} \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & -1 \end{pmatrix}\right) = -0.1\begin{pmatrix} -1.66619\times10^{-5} & 0 & 0 & 1.66619\times10^{-5} \\ -1.97816\times10^{-5} & 0 & 0 & 1.97816\times10^{-5} \\ -1.66619\times10^{-5} & 0 & 0 & 1.66619\times10^{-5} \end{pmatrix} = \begin{pmatrix} -1.66619\times10^{-6} & 0 & 0 & 1.66619\times10^{-6} \\ -1.97816\times10^{-6} & 0 & 0 & 1.97816\times10^{-6} \\ -1.66619\times10^{-6} & 0 & 0 & 1.66619\times10^{-6} \end{pmatrix}$$

$$\bullet\ \Delta b^{[1]} = -0.1\,\delta^{[1]} = -0.1\begin{pmatrix} -1.66619\times10^{-5} \\ -1.97816\times10^{-5} \\ -1.66619\times10^{-5} \end{pmatrix} = \begin{pmatrix} 1.66619\times10^{-6} \\ 1.97816\times10^{-6} \\ 1.66619\times10^{-6} \end{pmatrix}$$

$$\bullet\ \Delta W^{[2]} = -0.1\left(\delta^{[2]} \cdot (x^{[1]})^T\right) = -0.1\left(\begin{pmatrix} -1.1509\times10^{-5} \\ -1.72899\times10^{-4} \end{pmatrix}\begin{pmatrix} -0.90515 & -0.90515 & -0.90515 \end{pmatrix}\right) =$$

$$= -0.1\begin{pmatrix} 1.04176\times10^{-5} & 1.04176\times10^{-5} & 1.04176\times10^{-5} \\ 1.56499\times10^{-4} & 1.56499\times10^{-4} & 1.56499\times10^{-4} \end{pmatrix} = \begin{pmatrix} -1.04176\times10^{-6} & -1.04176\times10^{-6} & -1.04176\times10^{-6} \\ -1.56499\times10^{-5} & -1.56499\times10^{-5} & -1.56499\times10^{-5} \end{pmatrix}$$

$$\bullet\ \Delta b^{[2]} = -0.1\,\delta^{[2]} = -0.1\begin{pmatrix} -1.1509\times10^{-5} \\ -1.72899\times10^{-4} \end{pmatrix} = \begin{pmatrix} 1.1509\times10^{-6} \\ 1.72899\times10^{-5} \end{pmatrix}$$

$$\bullet\ \Delta W^{[3]} = -0.1\left(\delta^{[3]} \cdot (x^{[2]})^T\right) = -0.1\left(\begin{pmatrix} -0.0266 \\ 3.370\times10^{-6} \\ 0.00018 \end{pmatrix}\begin{pmatrix} -0.99956 & -0.99343 \end{pmatrix}\right) = -0.1\begin{pmatrix} 2.65845\times10^{-2} & 2.64215\times10^{-2} \\ -3.36815\times10^{-6} & -3.34749\times10^{-6} \\ -1.80384\times10^{-4} & -1.7928\times10^{-4} \end{pmatrix} = \begin{pmatrix} -2.65845\times10^{-3} & -2.64215\times10^{-3} \\ 3.36815\times10^{-7} & 3.34749\times10^{-7} \\ 1.80384\times10^{-5} & 1.7928\times10^{-5} \end{pmatrix}$$

$$\bullet\ \Delta b^{[3]} = -0.1\,\delta^{[3]} = -0.1\begin{pmatrix} -0.0266 \\ 3.370\times10^{-6} \\ 0.00018 \end{pmatrix} = \begin{pmatrix} 2.65961\times10^{-3} \\ -3.3696\times10^{-7} \\ -1.804\times10^{-5} \end{pmatrix}$$

Agora, já podemos atualizar os pesos e os biases:

$$\bullet\ W^{[3]}_{new} = W^{[3]}_{old} + \sum_{i=1}^{2}\Delta W_i^{[3]} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{pmatrix} + \left(\begin{pmatrix} -0.00031 & 0.00039 \\ 0.01431 & -0.01832 \\ -0.00031 & 0.00039 \end{pmatrix} + \begin{pmatrix} -2.65845\times10^{-3} & -2.64215\times10^{-3} \\ 3.36815\times10^{-7} & 3.34749\times10^{-7} \\ 1.80334\times10^{-5} & 1.7928\times10^{-5} \end{pmatrix}\right) =$$

$$= \begin{pmatrix} 0.99704 & 0.99775 \\ 3.01431 & 0.98169 \\ 0.99971 & 1.00041 \end{pmatrix}$$

$$\bullet\ b^{[3]}_{new} = b^{[3]}_{old} + \sum_{i=1}^{2}\Delta b_i^{[3]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \left(\begin{pmatrix} -0.00068 \\ 0.03177 \\ -0.00068 \end{pmatrix} + \begin{pmatrix} 2.65961\times10^{-3} \\ -3.3696\times10^{-7} \\ -1.804\times10^{-5} \end{pmatrix}\right) = \begin{pmatrix} 1.00198 \\ 1.03177 \\ 0.9993 \end{pmatrix}$$

$$\bullet\ W^{[2]}_{new} = W^{[2]}_{old} + \sum_{i=1}^{2}\Delta W_i^{[2]} = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \left(\begin{pmatrix} 0.01731 & 0.02852 & 0.01731 \\ 0.00469 & 0.00774 & 0.00469 \end{pmatrix} + \begin{pmatrix} -1.04176\times10^{-6} & -1.04176\times10^{-6} & -1.04176\times10^{-6} \\ -1.56499\times10^{-5} & -1.56499\times10^{-5} & -1.56499\times10^{-5} \end{pmatrix}\right) = \begin{pmatrix} 1.0173 & 4.02852 & 1.0173 \\ 1.00468 & 1.00772 & 1.00468 \end{pmatrix}$$

$$\bullet\ b^{[2]}_{new} = b^{[2]}_{old} + \sum_{i=1}^{2}\Delta b_i^{[2]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \left(\begin{pmatrix} 0.03745 \\ 0.01016 \end{pmatrix} + \begin{pmatrix} 1.1509\times10^{-6} \\ 1.72899\times10^{-5} \end{pmatrix}\right) = \begin{pmatrix} 1.03745 \\ 1.01017 \end{pmatrix}$$

$$\bullet\ W^{[1]}_{new} = W^{[1]}_{old} + \sum_{i=1}^{2}\Delta W_i^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} + \left(\begin{pmatrix} 0.01872 & 0.01872 & 0.01872 & 0.01872 \\ 0.03359 & 0.03359 & 0.03359 & 0.03359 \\ 0.01872 & 0.01872 & 0.01872 & 0.01872 \end{pmatrix} + \right.$$

$$\left. + \begin{pmatrix} -1.66619\times10^{-6} & 0 & 0 & 1.66619\times10^{-6} \\ -1.97816\times10^{-6} & 0 & 0 & 1.97816\times10^{-6} \\ -1.66619\times10^{-6} & 0 & 0 & 1.66619\times10^{-6} \end{pmatrix}\right) = \begin{pmatrix} 1.01872 & 1.01872 & 1.01872 & 1.01872 \\ 1.03359 & 1.03359 & 2.03359 & 1.03359 \\ 1.01872 & 1.01872 & 1.01872 & 1.01872 \end{pmatrix}$$

$$\bullet\ b^{[1]}_{new} = b^{[1]}_{old} + \sum_{i=1}^{2}\Delta b_i^{[1]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \left(\begin{pmatrix} 0.01872 \\ 0.03359 \\ 0.01872 \end{pmatrix} + \begin{pmatrix} 1.66619\times10^{-6} \\ 1.97816\times10^{-6} \\ 1.66619\times10^{-6} \end{pmatrix}\right) = \begin{pmatrix} 1.01872 \\ 1.03359 \\ 1.01872 \end{pmatrix}$$

Temos então que os pesos e os biases, após um batch gradient descent update, com learning rate $\eta = 0.1$, são:

- $$W^{[1]} = \begin{pmatrix} 1.01872 & 1.01872 & 1.01872 & 1.01872 \\ 1.03359 & 1.03359 & 2.93359 & 1.03359 \\ 1.01872 & 1.01872 & 1.01872 & 1.01872 \end{pmatrix} \qquad \bullet\ b^{[1]} = \begin{pmatrix} 1.01872 \\ 1.03359 \\ 1.01872 \end{pmatrix}$$

- $$W^{[2]} = \begin{pmatrix} 1.0173 & 4.02852 & 1.0173 \\ 1.00468 & 1.00772 & 1.00468 \end{pmatrix} \qquad \bullet\ b^{[2]} = \begin{pmatrix} 1.03745 \\ 1.01017 \end{pmatrix}$$

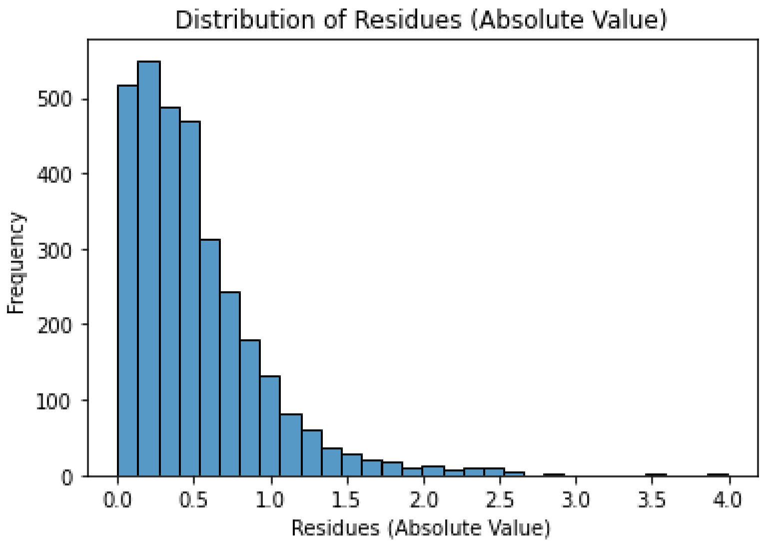
- $$W^{[3]} = \begin{pmatrix} 0.99704 & 0.99775 \\ 3.01431 & 0.98169 \\ 0.99971 & 1.00041 \end{pmatrix} \qquad \bullet\ b^{[3]} = \begin{pmatrix} 1.00198 \\ 1.03177 \\ 0.9993 \end{pmatrix}$$

II. Programming and critical analysis

1)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.neural_network import MLPRegressor
from sklearn.model_selection import train_test_split

# Reading the CSV file
df = pd.read_csv("winequality-red.csv", sep=";")

X = df.drop("quality", axis=1)
y = df["quality"]  # Get the "quality" column as the target variable

X_train, X_test, y_train, y_test = train_test_split(X.values, y, test_size=0.2, random_state=0)

# Residues
res = []

for state in range(1, 11):

  mlp = MLPRegressor(hidden_layer_sizes=(10, 10), activation='relu', random_state=state, 
early_stopping=True, validation_fraction=0.2)
  mlp.fit(X_train, y_train)

  # Predict and calculate the residues
  pred = mlp.predict(X_test)
    residue = np.abs(y_test - pred)
```

```
    res.extend(residue)

# Plot the histogram of the residues using Seaborn
sns.histplot(res, bins = 30)
plt.title("Distribution of Residues (Absolute Value)")
plt.xlabel("Residues (Absolute Value)")
plt.ylabel("Frequency")
plt.show()
```



2)

```
from sklearn.metrics import mean_absolute_error

# Initialize lists to store MAE before and after rounding and bounding
mae_original = []
mae_round = []

X_train, X_test, y_train, y_test = train_test_split(X.values, y, test_size=0.2, random_state=0)

# Loop through random seeds from 1 to 10
for state in range(1, 11):

    mlp = MLPRegressor(hidden_layer_sizes=(10,10), activation = 'relu', random_state=state,
            early_stopping = True, validation_fraction = 0.2)
    mlp.fit(X_train, y_train)

    y_pred = mlp.predict(X_test)
```

```
    # Calculate the MAE before rounding and bounding
    mae_original.append(mean_absolute_error(y_test, y_pred))


    # Round and Bound estimates
    round_pred = np.round(y_pred)
    y_rounded = np.clip(round_pred, 1, 10)

    # Calculate the MAE after rounding and bounding
    mae_round.append(mean_absolute_error(y_test, y_rounded))

# Calculate the average MAE before and after rounding and bounding
average_mae_original = np.mean(mae_original)
average_mae_round = np.mean(mae_round)

# Print the average MAE before and after
print(f"\nAverage MAE Before Round and Bound Estimates = {average_mae_original}")
print(f"Average MAE After Round and Bound Estimates = {average_mae_round}")
```

Average MAE Before Round and Bound Estimates = 0.5097171955009514

Average MAE After Round and Bound Estimates = 0.43875000000000003

3)

```
from sklearn.metrics import mean_squared_error
from math import sqrt

# MLP with Early Stopping
rmse_original = []

# Arrays for each max_iterations
rmse_20 = []
rmse_50 = []
rmse_100 = []
rmse_200 = []

rmse_arrays = [rmse_20, rmse_50, rmse_100, rmse_200]


iterations = [20, 50, 100, 200]

X_train, X_test, y_train, y_test = train_test_split(X.values, y, test_size=0.2, random_state=0)
```

```
# Early stopping MLP for RMSE comparison in exercise 4
for random_state in range(1, 11):

  original_mlp = MLPRegressor(hidden_layer_sizes=(10,10), activation='relu',
random_state=random_state,
                early_stopping=True, validation_fraction=0.2)
  original_mlp.fit(X_train, y_train)

  y_pred_original = original_mlp.predict(X_test)

  # RMSE
  rmse_original.append(sqrt(mean_squared_error(y_test, y_pred_original)))


# Max iterations
for num_iterations in iterations:
  for random_state in range(1, 11):

    mlp_iterations = MLPRegressor(hidden_layer_sizes=(10,10), activation='relu',
                random_state=random_state, max_iter=num_iterations)
    mlp_iterations.fit(X_train, y_train)


    y_pred_iterations = mlp_iterations.predict(X_test)

    # Calcula rmse
    rmse = sqrt(mean_squared_error(y_test, y_pred_iterations))

    if num_iterations == 20:
      rmse_20.append(rmse)
    elif num_iterations == 50:
      rmse_50.append(rmse)
    elif num_iterations == 100:
      rmse_100.append(rmse)
    elif num_iterations == 200:
      rmse_200.append(rmse)

means = []

# For better code, iterates through the array containing all the iterations
for i, rmse_array in enumerate(rmse_arrays):
  mean_value = np.mean(rmse_array)
    means.append((iterations[i], mean_value))
```

```
# Prints
for mean_rmse in means:
  print(f"Mean RMSE of {mean_rmse[0]} iterations: {mean_rmse[1]}")


mean_rmse_original = np.mean(rmse_original)

print(f"\nMean RMSE with Early Stopping = {mean_rmse_original}")
```

Mean RMSE of 20 iterations: 1.4039789509925442

Mean RMSE of 50 iterations: 0.7996073631460568

Mean RMSE of 100 iterations: 0.6940361469112144

Mean RMSE of 200 iterations: 0.6554543932216474

Mean RMSE with Early Stopping = 0.6706527958221328

4)

```
# Iterações do modelo com early stopping
# Usado para comparar os modelos
print(original_mlp.n_iter_)
```

200

R: Early stopping para o treino do modelo quando a performance em relação aos testes de validação começa a descer. Deste modo, este método é bom para evitar tanto o underfitting como o overfitting, treinando até ao momento que a accuracy para novos dados começa a descer. Com o número máximo de iterações podemos notar que o RMSE é maior para todos os modelos, exceto o que tem um máximo de 200 iterações. Podemos então concluir que nos outros modelos (20, 50 e 100 iterações) houve underfitting, não tendo iterações suficientes para treinar o modelo adequadamente (o que não acontece com o early stopping).

O early stopping separa os dados de treino em validation e training sets. Uma das suas desvantagens é o facto de ser sensível à escolha do validation set: caso o validation set seja muito pequeno ou não muito representativo do testing set, o modelo pode parar muito cedo ou muito tarde, levando a underfitting ou overfitting. Uma vez que os modelos com early stopping e max_iterations a 200 param nas 200 iterações (no caso do early stopping, porque é o max_iterations por default), e sendo o RMSE do modelo com early stopping maior, concluímos que o validation set não deve ser o mais representativo do testing set, levando a um erro maior.

END