Homework III – Group 108

(ist1103617, ist199991)

I. Pen-and-paper

1)

a)

Homework 3

1. Centers $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$ [illegible]

$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$:

$$\phi_j(x) = \exp\left(-\frac{\|(0,7;-0,3)-(0,0)\|^2}{2}\right) = 0,748$$

$$\phi_j(x) = \exp\left(-\frac{\|(0,4;0,5)-(0,0)\|^2}{2}\right) = 0,8146$$

$$\phi_j(x) = \exp\left(-\frac{\|(-0,2;0,8)-(0,0)\|^2}{2}\right) = 0,712$$

$$\phi_j(x) = \exp\left(-\frac{\|(-0,4;0,3)-(0,0)\|^2}{2}\right) = 0,8825$$

$\begin{pmatrix} 1 \\ -1 \end{pmatrix}$:

$$\phi_j(x) = \exp\left(-\frac{\|(0,7;-0,3)-(1;-1)\|^2}{2}\right) = 0,748$$

$$\phi_j(x) = \exp\left(-\frac{\|(0,4;0,5)-(1;-1)\|^2}{2}\right) = 0,271$$

$$\phi_j(x) = \exp\left(-\frac{\|(-0,2;0,8)-(1;-1)\|^2}{2}\right) = 0,096$$

$$\phi_j(x) = \exp\left(-\frac{\|(-0,4;0,3)-(1;-1)\|^2}{2}\right) = 0,161$$

$\begin{pmatrix} -1 \\ 1 \end{pmatrix}$:

$$\phi_j(x) = \exp\left(-\frac{\|(0,7;-0,3)-(-1;1)\|^2}{2}\right) = 0,101$$

$$\phi_j(x) = \exp\left(-\frac{\|(0,4;0,5)-(-1;1)\|^2}{2}\right) = 0,331$$

$$\phi_j(x) = \exp\left(-\frac{\|(-0,2;0,8)-(-1;1)\|^2}{2}\right) = 0,712$$

$$\phi_j(x) = \exp\left(-\frac{\|(-0,4;0,3)-(-1;1)\|^2}{2}\right) = 0,654$$

TÉCNICO LISBOA

Homework III – Group 108

(ist1103617, ist199991)

Minimize : $w^T w + \lambda w^T w$

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

Design Matrix

$$\phi = \begin{bmatrix} 1 & 0{,}748 & 0{,}748 & 0{,}101 \\ 1 & 0{,}8146 & 0{,}271 & 0{,}331 \\ 1 & 0{,}712 & 0{,}096 & 0{,}712 \\ 1 & 0{,}5825 & 0{,}161 & 0{,}654 \end{bmatrix}$$

$$\phi^T . \phi = \begin{bmatrix} 4 & 3{,}157 & 1{,}276 & 1{,}758 \\ 3{,}157 & 2{,}5 & 0{,}53 & 1{,}42 \\ 1{,}276 & 0{,}55 & 0{,}66 & 0{,}33 \\ 1{,}758 & 1{,}42 & 0{,}33 & 1{,}05 \end{bmatrix}$$

$$\phi^T . \phi + \lambda I = \begin{bmatrix} 4{,}1 & 3{,}157 & 1{,}276 & 1{,}758 \\ 3{,}157 & 2{,}6 & 0{,}55 & 1{,}42 \\ 1{,}276 & 0{,}55 & 0{,}768 & 0{,}33 \\ 1{,}758 & 1{,}42 & 0{,}33 & 1{,}15 \end{bmatrix}$$

$$(\phi^T . \phi + \lambda I)^{-1} = \begin{bmatrix} 4{,}55 & -3{,}38 & -1{,}86 & -1{,}85 \\ -3{,}38 & 3{,}58 & -0{,}88 & -1{,}26 \\ -1{,}86 & -0{,}88 & 4{,}33 & 2{,}21 \\ -1{,}85 & -1{,}26 & 2{,}72 & 4{,}52 \end{bmatrix}$$

$$\phi^T z = [2 \quad 1{,}565 \quad 0{,}8381 \quad 0{,}6852]$$

$$w = \begin{bmatrix} 0{,}339 \\ 0{,}199 \\ 0{,}4 \\ -0{,}296 \end{bmatrix} \; //$$

Bias

b)

Homework III – Group 108

(ist1103617, ist199991)

b)

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

$z_1 = 0{,}8$; $z_2 = 0{,}6$; $z_3 = 0{,}3$; $z_4 = 0{,}3$

$\hat{z} = w^T \phi(x)$

$\phi(x_i) = [1 \quad \phi_1(x_i) \quad \phi_2(x_i) \quad \phi_3(x_i)]^T$

$$\hat{z}_1 = [0{,}33914 \quad 0{,}15945 \quad 0{,}4 \quad -0{,}256] \cdot \begin{bmatrix} 1 \\ 0{,}748 \\ 0{,}748 \\ -0{,}1 \end{bmatrix} = 0{,}758$$

$$\hat{z}_2 = [0{,}33914 \quad 0{,}15945 \quad 0{,}4 \quad -0{,}256] \cdot \begin{bmatrix} 1 \\ 0{,}81 \\ 0{,}27 \\ 0{,}38 \end{bmatrix} = 0{,}51$$

$$\hat{z}_3 = [0{,}339 \quad 0{,}159 \quad 0{,}4 \quad -0{,}256] \cdot \begin{bmatrix} 1 \\ 0{,}71 \\ 0{,}03 \\ 0{,}71 \end{bmatrix} = 0{,}303$$

$$\hat{z}_4 = [0{,}339 \quad 0{,}159 \quad 0{,}4 \quad -0{,}256] \cdot \begin{bmatrix} 1 \\ 0{,}88 \\ 0{,}161 \\ 0{,}65 \end{bmatrix} = 0{,}386$$

$$RMSE = \frac{1}{2}\sqrt{(z_1-\hat{z}_1)^2 + (z_2-\hat{z}_2)^2 + (z_3-\hat{z}_3)^2 + (z_4-\hat{z}_4)^2} =$$

$$= 0{,}06503$$

2)

(2) Training Observations: $x_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$; $x_2 = \begin{pmatrix}1\\0\\0\\-1\end{pmatrix}$ Forward Propagation!

$f(x) = \tanh(0{,}5x - 2)$

x_1:

$$z[1] = \begin{pmatrix}1&1&1&1\\1&1&2&1\\1&1&1&1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}5\\6\\5\end{pmatrix}$$

$$f(z[1]) = \begin{pmatrix}0{,}46\\0{,}76\\0{,}46\end{pmatrix} - x^{(1)}_{x_1}$$

$$z[2] = \begin{pmatrix}1&4&1\\1&1&1\end{pmatrix}\begin{pmatrix}0{,}46\\0{,}76\\0{,}46\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}4{,}37\\2{,}68\end{pmatrix}$$

$$f(z[2]) = \begin{bmatrix}0{,}45\\-0{,}57\end{bmatrix} - x^{(2)}_{x_1}$$

$$z[3] = \begin{pmatrix}1&1\\3&1\\1&1\end{pmatrix}\cdot\begin{pmatrix}0{,}45\\-0{,}57\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0{,}874\\1{,}77\\0{,}874\end{pmatrix}$$

$$f(z[3]) = \begin{pmatrix}-0{,}51530\\-0{,}805\\-0{,}515\end{pmatrix} - x^{(3)}_{x_1}$$ B output $z = \begin{bmatrix}-1\\1\\-1\end{bmatrix}\begin{matrix}A\\B\\C\end{matrix}$

x_2:

$$z[1] = \begin{pmatrix}1&1&1&1\\1&1&2&1\\1&1&1&1\end{pmatrix}\cdot\begin{pmatrix}1\\0\\0\\-1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$f(z[1]) = \begin{pmatrix}-0{,}5515\\-0{,}5515\\-0{,}5515\end{pmatrix} - x^{(1)}_{x_2}$$

$$z[2] = \begin{pmatrix}1&4&1\\1&1&1\end{pmatrix}\cdot\begin{pmatrix}-0{,}5515\\-0{,}5515\\-0{,}5515\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-4{,}43\\-1{,}715\end{pmatrix}$$

$$f(z[2]) = \begin{pmatrix}-0{,}555\\-0{,}553\end{pmatrix} - x_1^{(2)}$$

$$z[3] = \begin{pmatrix}1&1\\3&1\\1&1\end{pmatrix}\cdot\begin{pmatrix}-0{,}555\\-0{,}553\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}-0{,}553\\-2{,}95\\-0{,}05\end{pmatrix}$$

$$f(z[3]) = \begin{pmatrix}-0{,}5965\\-0{,}558\\-0{,}986\end{pmatrix} - x^{(3)}_{x_2}$$ A output $z = \begin{bmatrix}1\\-1\\-1\end{bmatrix}\begin{matrix}A\\B\\C\end{matrix}$

$$\text{Activation\_Derivative}(x) = \frac{\partial \tanh(0,5x - 2)}{\partial x} =$$

$$= 0,5\,(1 - \tanh(0,5x - 2))$$

<u>Gradient of the error function</u>

$$\frac{\partial E(z, \hat{z})}{\partial W^{[L]}} = \delta^{[L]} \cdot (X^{[L-1]})^T$$

<u>Last Layer - Output</u>

$$\delta[L] = \frac{\partial E(z, \hat{z})}{\partial \hat{z}} \odot \text{activation\_derivative}(z^{[L]})$$

$$\frac{\partial E(z, \hat{z})}{\partial \hat{z}} = \frac{\partial \left(\frac{1}{2} \| z - \hat{z} \|_2^2\right)}{\partial \hat{z}} = \hat{z} - z$$

z - A or B

$\hat{z}$ - output from network, $X_i^{[3]} = f(z[3])$ de x_1 ou x_2

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

<u>Other Layers</u>

$$\delta^{[L]} = (W^{[L+1]})^T \delta^{[L+1]} \odot \text{activation\_derivative}(z^{[L]})$$

<u>Calculating $\delta^{[L]}$</u>

<u>Output:</u>

$$\delta^{[3]} = (X^{[3]} - z) \odot \text{activ\_derivative}(z^{[3]})$$

<u>x_1:</u>

$$\delta_{x_1}^{[3]} = \left[\begin{pmatrix} -0,915 \\ -0,805 \\ -0,915 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}\right] \odot f'(z^{[3]}) = \begin{bmatrix} 0,0068 \\ -0,3177 \\ 0,0068 \end{bmatrix}$$

<u>x_2:</u>

$$\delta_{x_1}^{[3]} = \left[\begin{pmatrix} -0,9865 \\ -0,358 \\ -0,5865 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}\right] \odot f'(z^{[3]}) = \begin{bmatrix} -0,0266 \\ 0,00000337 \\ 0,00018 \end{bmatrix}$$

Other Layers:

$$\delta^{[2]} = (W^{[3]})^T \delta^{[3]} \odot \text{activation\_derivative}(z^{[2]})$$
$$\delta^{[1]} = (W^{[2]})^T \delta^{[2]} \odot \text{activation\_derivative}(z^{[1]})$$

x_1:
$$\delta^{[2]}_{x_1} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}^T \begin{pmatrix} 0{,}0068 \\ -0{,}3177 \\ 0{,}0068 \end{pmatrix} \odot f'(z^{[2]}) = \begin{bmatrix} -0{,}3745 \\ -0{,}1016 \end{bmatrix}$$

$$\delta^{[1]}_{x_1} = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} -0{,}3745 \\ -0{,}1016 \end{pmatrix} \odot f'(z^{[1]}) = \begin{pmatrix} -0{,}1776 \\ -0{,}3335 \\ -0{,}1876 \end{pmatrix}$$

x_2:
$$\delta^{[2]}_{x_2} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}^T \begin{pmatrix} -0{,}0266 \\ 0{,}0000377 \\ 0{,}00018 \end{pmatrix} \odot f'(z^{[2]}) = \begin{bmatrix} 0{,}0000115 \\ 0{,}000173 \end{bmatrix}$$

$$\delta^{[1]}_{x_2} = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 0{,}0000115 \\ 0{,}000173 \end{pmatrix} \odot f'(z^{[1]}) = \begin{bmatrix} 0{,}00001666 \\ 0{,}00001978 \\ 0{,}00001666 \end{bmatrix}$$

$$\sum_{k=1}^{2} \frac{\partial E(t_k, \hat{t}_k)}{\partial W^{[3]}} = \delta^{[3]}_{x_1} (X^{[2]}_{x_1})^T + \delta^{[3]}_{x_2} (X^{[2]}_{x_2})^T =$$
$$= \begin{bmatrix} 0{,}0068 \\ -0{,}3177 \\ 0{,}0068 \end{bmatrix} [0{,}4505 \ \ -0{,}5764] + \begin{bmatrix} 0{,}026596 \\ 0{,}0000377 \\ 0{,}00018 \end{bmatrix} \circledast$$

$\circledast$
$$[-0{,}5556 \ \ -0{,}5534] =$$
$$= \begin{bmatrix} 0{,}0296 & 0{,}0225 \\ -0{,}1431 & 0{,}1831 \\ 0{,}0025 & -0{,}0041 \end{bmatrix} //$$

$$\sum_{k=1}^{2} \frac{\partial E(t_k, \hat{t}_k)}{\partial W^{[2]}} = \delta^{[2]}_{x_1} (X^{[1]}_{x_1})^T + \delta^{[2]}_{x_2} (X^{[1]}_{x_2})^T =$$
$$= \begin{pmatrix} -0{,}3745 \\ -0{,}1016 \end{pmatrix} \begin{pmatrix} 0{,}46 \\ 0{,}76 \\ 0{,}46 \end{pmatrix}^T + \begin{pmatrix} 0{,}0000115 \\ 0{,}000173 \end{pmatrix} \begin{pmatrix} -0{,}5515 \\ -0{,}5515 \\ -0{,}5515 \end{pmatrix} =$$
$$= \begin{bmatrix} -0{,}173 & -0{,}285 & -0{,}173 \\ -0{,}0467 & -0{,}077 & -0{,}0467 \end{bmatrix}$$

$$\sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{w^{[1]}} = \delta_{\times 1}^{[1]} (X_{\times 1}^{[0]})^T + \delta_{\times 2}^{[1]} (X_{\times 2}^{[0]})^T =$$

$$= \begin{pmatrix} -0{,}18772 \\ -0{,}33159 \\ -0{,}18772 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}^T + \begin{pmatrix} 0{,}0000\,1666 \\ 0{,}0000\,1578 \\ 0{,}0000\,1666 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} =$$

$$= \begin{bmatrix} -0{,}187 & -0{,}187 & -0{,}187 & -0{,}187 \\ -0{,}33 & -0{,}33 & -0{,}33 & -0{,}33 \\ -0{,}18 & -0{,}18 & -0{,}18 & -0{,}18 \end{bmatrix}$$

Updating weights:

$$w^{[1]} = w^{[1]} - \eta \sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial w^{[1]}} = \begin{bmatrix} 1{,}0187 & 1{,}0187 & 1{,}0187 & 1{,}0187 \\ 1{,}033 & 1{,}033 & 2{,}033 & 1{,}033 \\ 1{,}018 & 1{,}018 & 1{,}018 & 1{,}018 \end{bmatrix}$$

$$w^{[2]} = w^{[2]} - \eta \sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial w^{[2]}} = \begin{bmatrix} 1{,}0173 & 4{,}0285 & 1{,}0173 \\ 1{,}00467 & 1{,}0077 & 1{,}00467 \end{bmatrix}$$

$$w^{[3]} = w^{[3]} - \eta \sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial w^{[3]}} = \begin{bmatrix} 0{,}9570 & 0{,}5577 \\ 3{,}0743 & 0{,}9817 \\ 0{,}9557 & 1{,}000 \end{bmatrix}$$

Update Biases: Calculation

$$\frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[l]}} = \frac{\partial E(z_k, \hat{z}_k)}{\partial z^{[l]}} \times \frac{\partial z^{[l]}}{\partial b^{[l]}} \times 1 = \delta^{[l]}$$

$$\sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[1]}} = \delta^{[1]}_{x_1} + \delta^{[1]}_{x_2} = \begin{bmatrix} -0,1872 \\ -0,3355 \\ -0,1772 \end{bmatrix}.$$

$$\sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[2]}} = \delta^{[2]}_{x_1} + \delta^{[2]}_{x_2} = \begin{bmatrix} -0,3745 \\ -0,1017 \end{bmatrix}$$

$$\sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[3]}} = \delta^{[3]}_{x_1} + \delta^{[3]}_{x_2} = \begin{bmatrix} -0,0192 \\ -0,3177 \\ 0,0070 \end{bmatrix}$$

Update Biases:

$$b^{[1]'} = b^{[1]} - \eta \sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[1]}} = \begin{bmatrix} 1,0187 \\ 1,0336 \\ 1,0177 \end{bmatrix}$$

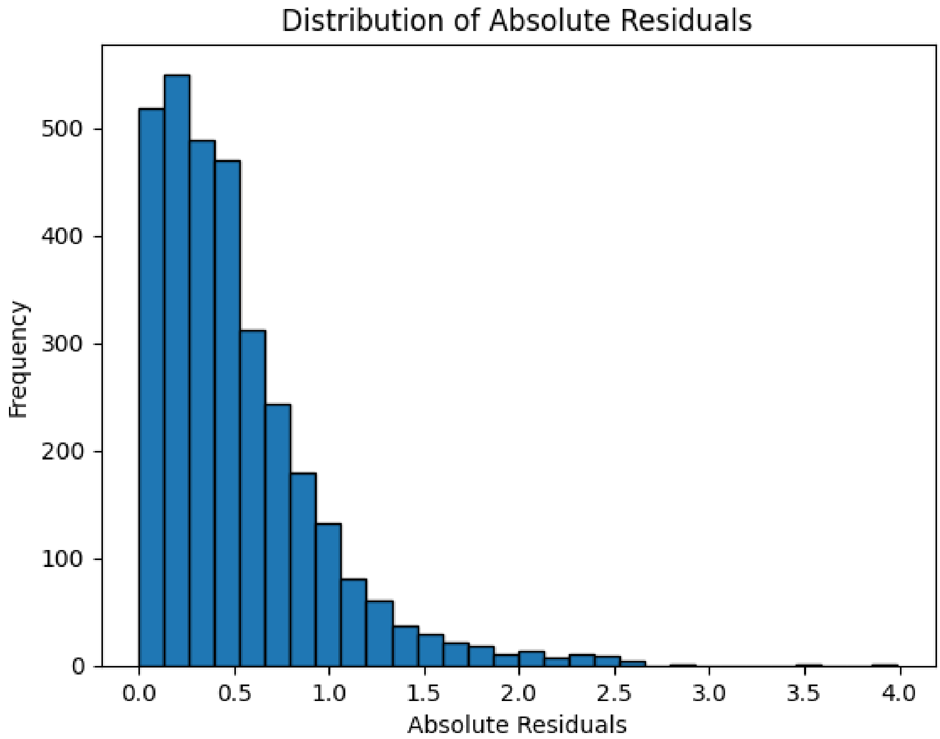
$$b^{[2]'} = b^{[2]} - \eta \sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[2]}} = \begin{bmatrix} 1,0374 \\ 1,0101 \end{bmatrix}$$

$$b^{[3]'} = b^{[3]} - \eta \sum_{k=1}^{2} \frac{\partial E(z_k, \hat{z}_k)}{\partial b^{[3]}} = \begin{bmatrix} 1,0020 \\ 1,0318 \\ 0,9553 \end{bmatrix}$$

Homework III – Group 108
(ist1103617, ist199991)

II. Programming and critical analysis

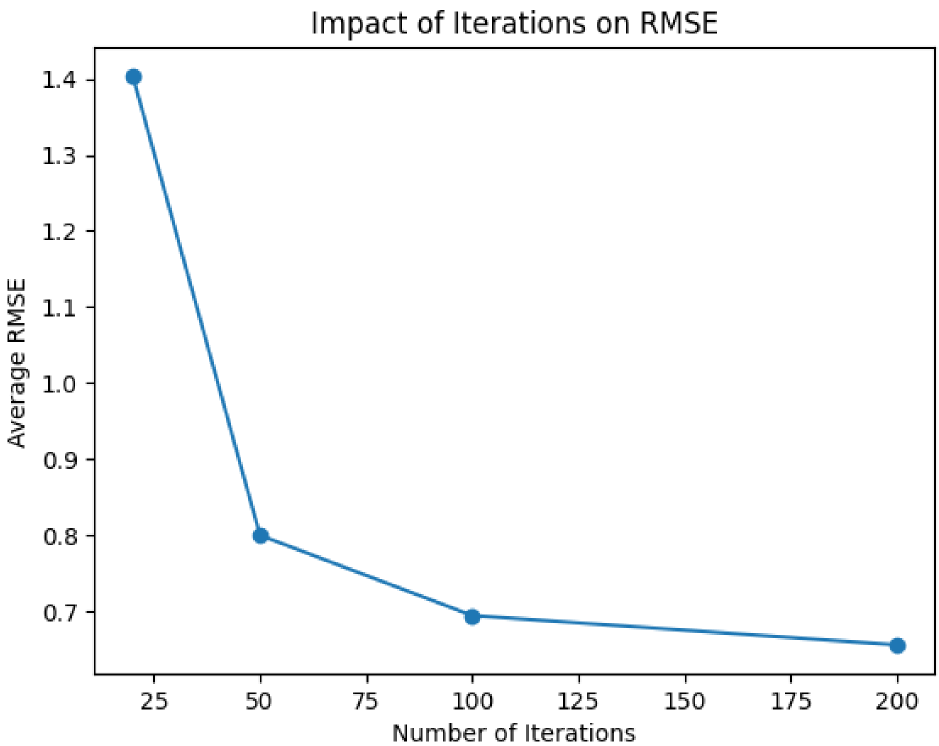
1)



2) Original MAE: 0.5097171955009514 Rounded and Bounded MAE: 0.43875. The second MAE value suggests that rounding and bounding are effective techniques for improving the model's performance in this specific regression task.

3) RMSE with 20 iterations: 1.403978950992544
RMSE with 50 iterations: 0.7996073631460567

RMSE with 100 iterations: 0.6940361469112146
RMSE with 200 iterations: 0.6554543932216473



4) Early stopping at significantly smaller iteration numbers drastically worsens performance, since the model does not have the opportunities to refine its predictions. On the other hand, stopping at 200 iterations, while not perfect, give us a much better RMSE value. Having said this, more is not always better. We should be careful in finding a balance between overfitting and underfitting in order to achieve the best RMSE values possible.


Homework3.ipynb

END