Aprendizagem 2022

Homework II – Group 58

**Part I**: Pen and paper

1.

1

applying the feature transformation of $\varphi(x_1)$:

$$\begin{array}{c} \begin{array}{cccc} x^0 & x^1 & x^2 & x^3 \end{array} \\ \begin{bmatrix} 1 & 0.8 & 0.8^2 & 0.8^3 \\ 1 & 1 & 1^2 & 1^3 \\ 1 & 1.2 & 1.2^2 & 1.2^3 \\ 1 & 1.4 & 1.4^2 & 1.4^3 \\ 1 & 1.6 & 1.6^2 & 1.6^3 \end{bmatrix} \end{array} = \begin{bmatrix} 1 & 0.8 & 0.64 & 0.512 \\ 1 & 1 & 1 & 1 \\ 1 & 1.2 & 1.44 & 1.728 \\ 1 & 1.4 & 1.96 & 2.744 \\ 1 & 1.6 & 2.56 & 4.096 \end{bmatrix} = X$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.8 & 1 & 1.2 & 1.4 & 1.6 \\ 0.64 & 1 & 1.44 & 1.96 & 2.56 \\ 0.512 & 1 & 1.728 & 2.744 & 4.096 \end{bmatrix}$$

$$X^T \cdot X = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.8 & 1 & 1.2 & 1.4 & 1.6 \\ 0.64 & 1 & 1.44 & 1.96 & 2.56 \\ 0.512 & 1 & 1.728 & 2.744 & 4.096 \end{bmatrix}}_{4\times 5} \underbrace{\begin{bmatrix} 1 & 0.8 & 0.64 & 0.512 \\ 1 & 1 & 1 & 1 \\ 1 & 1.2 & 1.44 & 1.728 \\ 1 & 1.4 & 1.96 & 2.744 \\ 1 & 1.6 & 2.56 & 4.096 \end{bmatrix}}_{5\times 4}$$

$$= \underbrace{\begin{bmatrix} 5 & 6 & 7.6 & 10.08 \\ 6 & 7.6 & 10.08 & 13.8784 \\ 7.6 & 10.08 & 13.8784 & 19.68 \\ 10.08 & 13.8784 & 19.68 & 28.55488 \end{bmatrix}}_{4\times 4} \qquad \lambda I = \underbrace{\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}}_{4\times 4}$$

$$X^T \cdot X + \lambda I = \begin{bmatrix} 7 & 6 & 7.6 & 10.08 \\ 6 & 9.6 & 10.08 & 13.8784 \\ 7.6 & 10.08 & 15.8784 & 19.68 \\ 10.08 & 13.8784 & 19.68 & 30.55488 \end{bmatrix}$$

$$(X^T \cdot X + \lambda I)^{-1} = \begin{bmatrix} 0.3417 & -0.1214 & -0.0749 & -0.0093 \\ -0.1214 & 0.3892 & -0.0967 & -0.0745 \\ -0.0749 & -0.0967 & 0.3726 & -0.1714 \\ -0.0093 & -0.0745 & -0.1714 & 0.1800 \end{bmatrix}$$

$$W = (X^T \cdot X + \lambda I)^{-1} X^T z$$

$$(X^T \cdot X + \lambda I)^{-1} X^T = \underbrace{\begin{bmatrix} 0.3417 & -0.1214 & -0.0749 & -0.0093 \\ -0.1214 & 0.3892 & -0.0967 & -0.0745 \\ -0.0749 & -0.0967 & 0.3726 & -0.1714 \\ -0.0093 & -0.0745 & -0.1714 & 0.1800 \end{bmatrix}}_{4 \times 4} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.8 & 1 & 1.2 & 1.4 & 1.6 \\ 0.64 & 1 & 1.44 & 1.96 & 2.56 \\ 0.512 & 1 & 1.728 & 2.744 & 4.096 \end{bmatrix}}_{4 \times 5}$$

$$= \underbrace{\begin{bmatrix} 0.1918 & 0.1360 & 0.0720 & -0.0007 & -0.0825 \\ 0.0899 & 0.0966 & 0.0777 & 0.0297 & -0.0512 \\ -0.0015 & 0.0296 & 0.0495 & 0.0498 & 0.0224 \\ -0.0864 & -0.0751 & -0.0344 & 0.0445 & 0.1701 \end{bmatrix}}_{4 \times 5}$$

$$(X^T \cdot X + \lambda I)^{-1} X^T z =$$

$$\underbrace{\begin{bmatrix} 0.1918 & 0.1360 & 0.0720 & -0.0007 & -0.0825 \\ 0.0899 & 0.0966 & 0.0777 & 0.0297 & -0.0512 \\ -0.0015 & 0.0296 & 0.0495 & 0.0498 & 0.0224 \\ -0.0864 & -0.0751 & -0.0344 & 0.0445 & 0.1701 \end{bmatrix}}_{4 \times 5} \underbrace{\begin{bmatrix} 24 \\ 26 \\ 10 \\ 13 \\ 12 \end{bmatrix}}_{5 \times 1} = \begin{bmatrix} 7.0451 \\ 4.6409 \\ 1.9673 \\ -1.3009 \end{bmatrix}$$

$$\hat{y}(x) = \text{output}(x) = W^t \phi(x) = \begin{bmatrix} 7.0451 & 4.6409 & 1.9673 & -1.3009 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} =$$

$$= 7{,}0451 + 4{,}6409\,x + 1.9673x^2 - 1.3009x^3$$

2.

(2)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad \hat{y}(x) = 7{,}0451 + 4{,}6409x + 1.9673x^2 - 1.3009x^3$$

$$\hat{y}(0.8) = 11.3508 \qquad \hat{y}(1.4) = 13.8286$$

$$\hat{y}(1) = 12.3524 \qquad \hat{y}(1.6) = 14.1783$$

$$\hat{y}(1.2) = 13.19914$$

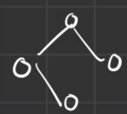
$$RMSE = \sqrt{\frac{(11.3508 - 24)^2 + (12.3524 - 20)^2 + (13.19914 - 10)^2 + (13.8286 - 13)^2 + (14.1783 - 12)^2}{5}}$$

$$= \sqrt{\frac{160.0023 + 58.4858 + 10.2345 + 0.6866 + 2.1783}{5}} = \sqrt{\frac{231.5875}{5}}$$

$$= \sqrt{46.3175} = 6.8057$$

3.

3.



$u = [0{,}8], [1], [1{,}2]$

$G = \frac{1}{2} \sum_{i}^{m} (z_i - \hat{z}_i)^2$

$\frac{\partial G}{\partial \hat{z}} = \hat{z}_i - z_i$

$W_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$W_2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad b_2 = \begin{bmatrix} 1 \end{bmatrix}$

Forward

$h_1 = W_1 \cdot u + b_1$

$z_1 = e^{0{,}1 h_1}$

$h_2 = W_2 \cdot z_1 + b_2$

$\hat{z} = e^{0{,}1 h_2}$

$u = 0{,}8$

$$h_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot 0{,}8 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1{,}8 \\ 1{,}8 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} e^{0{,}1 \cdot 1{,}8} \\ e^{0{,}1 \cdot 1{,}8} \end{bmatrix} = \begin{bmatrix} 1{,}197 \\ 1{,}197 \end{bmatrix}$$

$$h_2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1{,}197 \\ 1{,}197 \end{bmatrix} + 1 = 3{,}394$$

$$\hat{z} = e^{0{,}1 \cdot 3{,}394} = 1{,}404$$

$u = 1$

$$h_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot 1 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} e^{0{,}1 \cdot 2} \\ e^{0{,}1 \cdot 2} \end{bmatrix} = \begin{bmatrix} 1{,}221 \\ 1{,}221 \end{bmatrix}$$

$$h_2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1{,}221 \\ 1{,}221 \end{bmatrix} + 1 = 3{,}442$$

$$\hat{z} = e^{0{,}1 \cdot 3{,}442} = 1{,}411$$

$u = 1{,}2$

$$h_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot 1{,}2 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2{,}2 \\ 2{,}2 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} e^{0{,}1 \cdot 2{,}2} \\ e^{0{,}1 \cdot 2{,}2} \end{bmatrix} = \begin{bmatrix} 1{,}246 \\ 1{,}246 \end{bmatrix}$$

$$h_2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1{,}246 \\ 1{,}246 \end{bmatrix} + 1 = 3{,}492$$

$$\hat{z} = e^{0{,}1 \cdot 3{,}492} = 1{,}418$$

Backward

$$b_2 = b_2 - \eta \frac{\partial G}{\partial b_2}$$

$$w_2 = w_2 - \eta \frac{\partial G}{\partial w_2}$$

$$\frac{\partial G}{\partial \hat{z}} = \hat{z} - z \qquad \frac{\partial h_2}{\partial b_2} = 1$$

$$\frac{\partial \hat{z}}{\partial h_2} = 0{,}1\, e^{0{,}1 h_2}$$

$$\frac{\partial G}{\partial b_2} = \left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right) \otimes \frac{\partial h_2}{\partial b_2}$$

$$\frac{\partial h_2}{\partial w_2} = z_1$$

$$\frac{\partial G}{\partial w_2} = \left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right) \otimes \frac{\partial h_2}{\partial w_2}$$

$u = 0{,}8$

$$\frac{\partial G}{\partial \hat{z}} = 1{,}404 - 24 = -22{,}596 \qquad \frac{\partial \hat{z}}{\partial h_2} = 0{,}1\, e^{0{,}1 \cdot 3{,}394} = 0{,}140 \qquad \frac{\partial h_2}{\partial w_2} = \begin{bmatrix} 1{,}197 \\ 1{,}197 \end{bmatrix}$$

$u = 1$

$$\frac{\partial G}{\partial \hat{z}} = 1{,}411 - 20 = -18{,}589 \qquad \frac{\partial \hat{z}}{\partial h_2} = 0{,}1\, e^{0{,}1 \cdot 3{,}442} = 0{,}141 \qquad \frac{\partial h_2}{\partial w_2} = \begin{bmatrix} 1{,}221 \\ 1{,}221 \end{bmatrix}$$

$u = 1{,}2$

$$\frac{\partial G}{\partial \hat{z}} = 1{,}418 - 10 = -8{,}582 \qquad \frac{\partial \hat{z}}{\partial h_2} = 0{,}1\, e^{0{,}1 \cdot 3{,}492} = 0{,}142 \qquad \frac{\partial h_2}{\partial w_2} = \begin{bmatrix} 1{,}246 \\ 1{,}246 \end{bmatrix}$$

$$b_2 = 1 - 0{,}1\left(\left((-22{,}596 \odot 0{,}140) \otimes 1\right) + \left((-18{,}589 \odot 0{,}141) \otimes 1\right) + \left((-8{,}582 \odot 0{,}142) \otimes 1\right)\right)$$
$$= 1{,}7003$$

$$w_2 = \begin{bmatrix} 1 & 1 \end{bmatrix} - 0{,}1\left(\left(-3{,}163 \otimes \begin{bmatrix} 1{,}197 \\ 1{,}197 \end{bmatrix}\right) + \left(-2{,}621 \otimes \begin{bmatrix} 1{,}221 \\ 1{,}221 \end{bmatrix}\right) + \left(-1{,}219 \otimes \begin{bmatrix} 1{,}246 \\ 1{,}246 \end{bmatrix}\right)\right)$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} - 0{,}1\left(\begin{bmatrix} -3{,}786 & -3{,}786 \end{bmatrix} + \begin{bmatrix} -3{,}2 & -3{,}2 \end{bmatrix} + \begin{bmatrix} -1{,}519 & -1{,}519 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} - 0{,}1\left(\begin{bmatrix} -8{,}505 & -8{,}505 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \end{bmatrix} + \begin{bmatrix} 0{,}8505 & 0{,}8505 \end{bmatrix}$$
$$= \begin{bmatrix} 1{,}8505 & 1{,}8505 \end{bmatrix}$$

## Backward

$$b_1 = b_1 - n\frac{\partial G}{\partial b_1}$$

$$b_1 = b_1 - n\left(\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial b_1}\right)$$

$$w_1 = w_1 - n\frac{\partial G}{\partial w_1}$$

$$w_1 = w_1 - n\left(\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial w_1}\right)$$

$$\frac{\partial h_2}{\partial z_1} = w_2$$

$$\frac{\partial z_1}{\partial h_1} = 0{,}1e^{0{,}1h_1}$$

$$\frac{\partial h_1}{\partial b_1} = 1 \qquad \frac{\partial h_1}{\partial w_1} = u$$

## $u = 0{,}8$

$$\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T = -3{,}163 \qquad \frac{\partial h_2}{\partial z_1} = w_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T = \begin{bmatrix} -3{,}163 \\ -3{,}163 \end{bmatrix} \qquad \frac{\partial z_1}{\partial h_1} = 0{,}1 \cdot \begin{bmatrix} e^{0{,}1 \cdot 1{,}8} \\ e^{0{,}1 \cdot 1{,}8} \end{bmatrix} = \begin{bmatrix} 0{,}120 \\ 0{,}120 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} = \begin{bmatrix} -3{,}163 \\ -3{,}163 \end{bmatrix} \odot \begin{bmatrix} 0{,}120 \\ 0{,}120 \end{bmatrix} = \begin{bmatrix} -0{,}380 \\ -0{,}380 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial b_1} = \begin{bmatrix} -0{,}380 \\ -0{,}380 \end{bmatrix} \otimes 1 = \begin{bmatrix} -0{,}380 \\ -0{,}380 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial w_1} = \begin{bmatrix} -0{,}380 \\ -0{,}380 \end{bmatrix} \otimes 0{,}8 = \begin{bmatrix} -0{,}304 \\ -0{,}304 \end{bmatrix}$$

$u = 1$

$$\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T = -2{,}621 \qquad \frac{\partial h_2}{\partial z_1} = W_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T = \begin{bmatrix} -2{,}621 \\ -2{,}621 \end{bmatrix} \qquad \frac{\partial z_1}{\partial h_1} = 0{,}1 \cdot \begin{bmatrix} e^{0{,}1 \cdot 2} \\ e^{0{,}1 \cdot 2} \end{bmatrix} = \begin{bmatrix} 0{,}122 \\ 0{,}122 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} = \begin{bmatrix} -2{,}621 \\ -2{,}621 \end{bmatrix} \odot \begin{bmatrix} 0{,}122 \\ 0{,}122 \end{bmatrix} = \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial b_1} = \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix} \otimes 1 = \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial w_1} = \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix} \otimes 1 = \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix}$$

$u = 1{,}2$

$$\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T = -1{,}219 \qquad \frac{\partial h_2}{\partial z_1} = W_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T = \begin{bmatrix} -1{,}219 \\ -1{,}219 \end{bmatrix} \qquad \frac{\partial z_1}{\partial h_1} = 0{,}1 \cdot \begin{bmatrix} e^{0{,}1 \cdot 2{,}2} \\ e^{0{,}1 \cdot 2{,}2} \end{bmatrix} = \begin{bmatrix} 0{,}125 \\ 0{,}125 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} = \begin{bmatrix} -1{,}219 \\ -1{,}219 \end{bmatrix} \odot \begin{bmatrix} 0{,}125 \\ 0{,}125 \end{bmatrix} = \begin{bmatrix} -0{,}152 \\ -0{,}152 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial b_1} = \begin{bmatrix} -0{,}152 \\ -0{,}152 \end{bmatrix} \otimes 1 = \begin{bmatrix} -0{,}152 \\ -0{,}152 \end{bmatrix}$$

$$\left(\left(\frac{\partial G}{\partial \hat{z}} \odot \frac{\partial \hat{z}}{\partial h_2}\right)^T \cdot \frac{\partial h_2}{\partial z_1}\right)^T \odot \frac{\partial z_1}{\partial h_1} \otimes \frac{\partial h_1}{\partial w_1} = \begin{bmatrix} -0{,}152 \\ -0{,}152 \end{bmatrix} \otimes 1{,}2 = \begin{bmatrix} -0{,}1824 \\ -0{,}1824 \end{bmatrix}$$

$$b_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \left( \begin{bmatrix} -0{,}380 \\ -0{,}380 \end{bmatrix} + \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix} + \begin{bmatrix} -0{,}152 \\ -0{,}152 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0{,}0852 \\ 0{,}0852 \end{bmatrix} = \begin{bmatrix} 1{,}0852 \\ 1{,}0852 \end{bmatrix}$$

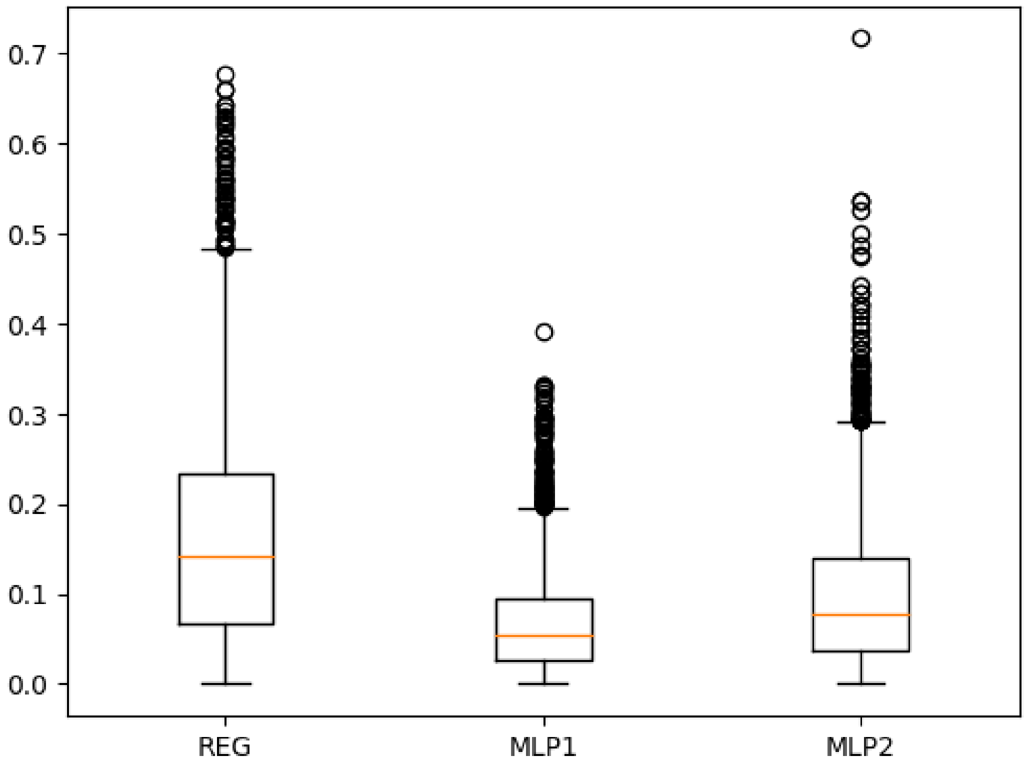
$$w_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \left( \begin{bmatrix} -0{,}304 \\ -0{,}304 \end{bmatrix} + \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix} + \begin{bmatrix} -0{,}1824 \\ -0{,}1824 \end{bmatrix} \right)$$

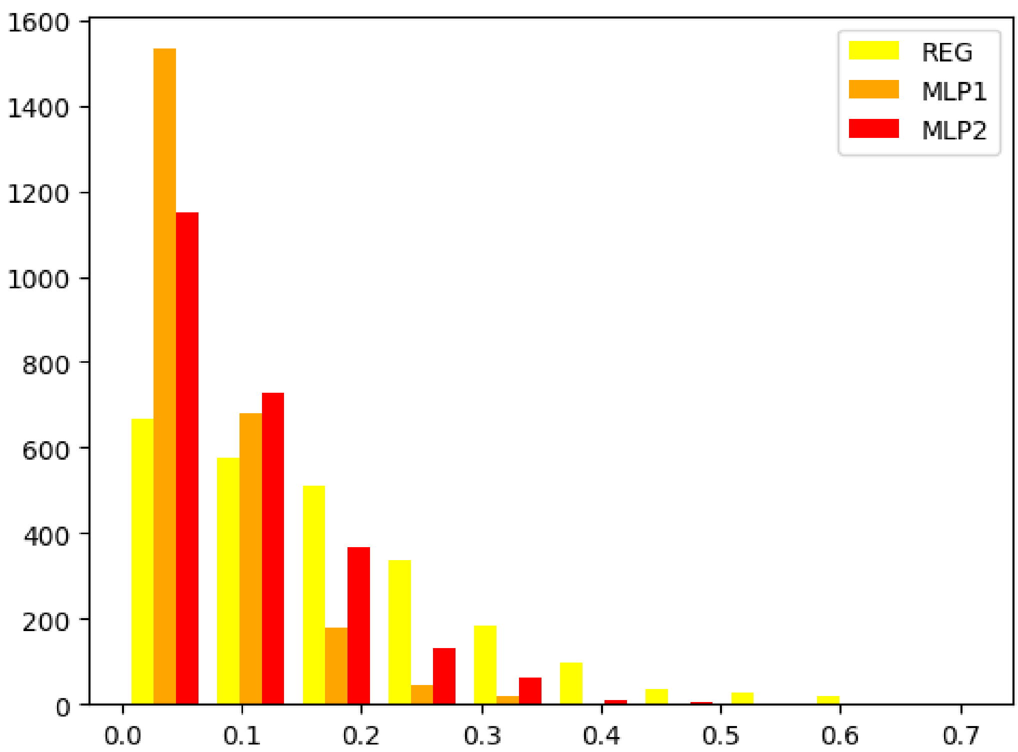
$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0{,}0804 \\ 0{,}0804 \end{bmatrix} = \begin{bmatrix} 1{,}0804 \\ 1{,}0804 \end{bmatrix}$$

**Part II**: Programming

4. $Ridge$ MAE: 0.162829976437694
   $MLP_1$ MAE: 0.0680414073796843
   $MLP_2$ MAE: 0.0978071820387748

5.



6. $MLP_1$ number of iterations: 452
   $MLP_2$ number of iterations: 77

7. We can observe that the Mean Absolute Error of $MLP_1$ is lower thsn the MSE of $MLP_2$ and the number of iterations of $MLP1$ is significantly higher (about 6 times) than those of $MLP_2$.
   In $MLP_1$, early stopping was considered, which means the training is stopped if the MSE on the validation set becomes bigger, even if the MSE on the training set goes down. In $MPL_1$, the MSE on the validation set (10% of training data) of our problem only starts growing after 452 iterations, way after the condition to stop when $early\_stopping = false$ (which is when the loss or score is not improving for $n\_iter\_no\_change$ consecutive iterations) is met.

# Appendix

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.io.arff import loadarff
from sklearn.linear_model import Ridge
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error
from sklearn.neural_network import MLPRegressor

data = loadarff('kin8nm.arff')
df = pd.DataFrame(data[0])
X = df.drop('y', axis=1)
y = df['y']

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.7,
    random_state = 0)

reg = Ridge(alpha = 0.1)
reg.fit(X_train, y_train)
y_pred_reg = reg.predict(X_test)
print('Ridge MAE:', mean_absolute_error(y_test, y_pred_reg))

mlp1 = MLPRegressor(hidden_layer_sizes = (10, 10,), activation = 'tanh', max_iter =
     500, random_state = 0, early_stopping = True)
mlp1.fit(X_train.values, y_train.values)
y_pred_mlp1 = mlp1.predict(X_test.values)
print('MLP1 MAE:', mean_absolute_error(y_test, y_pred_mlp1))

mlp2 = MLPRegressor(hidden_layer_sizes = (10, 10,), activation = 'tanh', max_iter =
     500, random_state = 0, early_stopping = False)
mlp2.fit(X_train.values, y_train.values)
y_pred_mlp2 = mlp2.predict(X_test.values)
print('MLP2 MAE:', mean_absolute_error(y_test, y_pred_mlp2))

residue_reg = np.array(abs(y_pred_reg - y_test))
residue_mlp1 = np.array(abs(y_pred_mlp1 - y_test))
residue_mlp2 = np.array(abs(y_pred_mlp2 - y_test))

fig, ax = plt.subplots()

ax.boxplot([residue_reg, residue_mlp1, residue_mlp2])
ax.set_xticklabels(['REG', 'MLP1','MLP2'])
plt.show()

plt.hist([residue_reg, residue_mlp1, residue_mlp2], 10,color=['yellow', 'orange', '
    red'], label=['REG', 'MLP1', 'MLP2'])
plt.legend()
plt.show()

print('MLP1 num of iter', mlp1.n_iter_)
print('MLP2 num of iter', mlp2.n_iter_)
```