Aprendizagem 2022/23

Homework III – Group 102

I. Pen-and-paper

1)

$$Z = \begin{bmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}8^2 & 1^2 & 1{,}2^2 & 1{,}4^2 & 1{,}6^2 \\ 0{,}8^3 & 1^3 & 1{,}2^3 & 1{,}4^3 & 1{,}6^3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{bmatrix}$$

$\lambda = 2$

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

$$XX^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 1{,}2 & 1{,}4 & 1{,}6 \\ 0{,}64 & 1 & 1{,}44 & 1{,}96 & 2{,}56 \\ 0{,}512 & 1 & 1{,}728 & 2{,}744 & 4{,}096 \end{bmatrix} \begin{bmatrix} 1 & 0{,}8 & 0{,}64 & 0{,}512 \\ 1 & 1 & 1 & 1 \\ 1 & 1{,}2 & 1{,}44 & 1{,}728 \\ 1 & 1{,}4 & 1{,}96 & 2{,}744 \\ 1 & 1{,}6 & 2{,}56 & 4{,}096 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 6 & 7{,}6 & 10{,}08 \\ 6 & 7{,}6 & 10{,}08 & 13{,}878 \\ 7{,}6 & 10{,}08 & 13{,}878 & 19{,}68 \\ 10{,}08 & 13{,}878 & 19{,}68 & 28{,}555 \end{bmatrix}$$

$$XX^T + \lambda I = \begin{bmatrix} 7 & 6 & 7,6 & 10,08 \\ 6 & 9,6 & 10,08 & 13,878 \\ 7,6 & 10,08 & 15,878 & 19,68 \\ 10,08 & 13,878 & 19,68 & 30,555 \end{bmatrix}$$

$$(XX^T + \lambda I)^{-1} XZ = \begin{bmatrix} 7,045 & 4,641 & 1,967 & -1,301 \end{bmatrix}$$

2) $\hat{z}_1 = 7,045 + 4,641 \times 0,8 + 1,967 \times 0,8^2 - 1,301 \times 0,8^3$
$= 11,351$

$\hat{z}_2 = 7,045 + 4,641 \times 1 + 1,967 \times 1^2 - 1,301 \times 1^3$
$= 12,352$

$\hat{z}_3 = 7,045 + 4,641 \times 1,2 + 1,967 \times 1,2^2 - 1,301 \times 1,2^3$
$= 13,199$

$\hat{z}_4 = 7,045 + 4,641 \times 1,4 + 1,967 \times 1,4^2 - 1,301 \times 1,4^3$
$= 13,828$

$\hat{z}_5 = 7,045 + 4,641 \times 1,6 + 1,967 \times 1,6^2 - 1,301 \times 1,6^3$
$= 14,177$

$$RMSE = \sqrt{\frac{\Sigma(\hat{z}_i - z_i)^2}{5}} = 6,843$$

3) $f(x) = e^{0,1x}$ $\quad NET^{[1]} = W^{[1]}x + b^{[1]}$

$x = 0,8$:

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

$$a^{[1]} = \begin{bmatrix} e^{0,1\times 1,8} \\ e^{0,1\times 1,8} \end{bmatrix} = \begin{bmatrix} e^{0,18} \\ e^{0,18} \end{bmatrix} = \begin{bmatrix} 1,197 \\ 1,197 \end{bmatrix}$$

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

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1,197 \\ 1,197 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3,394 \end{bmatrix}$$

$$a^{[2]} = \begin{bmatrix} e^{0,1\times 3,394} \end{bmatrix} = \begin{bmatrix} e^{0,3394} \end{bmatrix} =$$

$$= \begin{bmatrix} 1,404 \end{bmatrix}$$

$x = 1$:

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

$$a^{[1]} = \begin{bmatrix} e^{0,1\times 2} \\ e^{0,1\times 2} \end{bmatrix} = \begin{bmatrix} e^{0,2} \\ e^{0,2} \end{bmatrix} = \begin{bmatrix} 1,221 \\ 1,221 \end{bmatrix}$$

$$NET^{[2]} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1,221 \\ 1,221 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} 3,442 \end{bmatrix}$$

$$a^{[2]} = \begin{bmatrix} e^{3,442\times 0,1} \\ e^{3,442\times 0,1} \end{bmatrix} = \begin{bmatrix} e^{0,3442} \\ e^{0,3442} \end{bmatrix} = \begin{bmatrix} 1,411 \end{bmatrix}$$

$x = 1,2;$

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

$$a^{[1]} = \begin{bmatrix} e^{0,1 \times 2,2} \\ e^{0,1 \times 2,2} \end{bmatrix} = \begin{bmatrix} 1,246 \\ 1,246 \end{bmatrix}$$

$$NET^{[2]} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1,246 \\ 1,246 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} 3,492 \end{bmatrix}$$

$$a^{[2]} = \begin{bmatrix} e^{0,1 \times 3,492} \end{bmatrix} = \begin{bmatrix} 1,418 \end{bmatrix}$$

$$X \rightarrow NET^{[1]} \rightarrow a^{[1]} \rightarrow NET^{[2]} \rightarrow a^{[2]} \rightarrow l$$

$$l = \frac{(\hat{z} - z)}{2} \qquad \frac{dl}{da^2} = a^2 - z \qquad \frac{da^{[2]}}{dNET^{[2]}} = 0,1e^{0,1NET^{[2]}} = 0,1a^{[2]}$$

$$\frac{dl}{db^{[2]}} = \delta^{[2]} = \frac{da^{[2]}}{dNET^{[2]}} \frac{dl}{da^{[2]}} = 0,1\, a^{[2]} (a^{[2]} - z)$$

$$\frac{dNET^{[2]}}{da^{[1]}} = (w^2)^T = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\frac{da^{[1]}}{dNET^{[1]}} = \begin{bmatrix} 0,1e^{0,1NET^{[1]}} & 0 \\ 0 & 0,1e^{0,1NET^{[1]}} \end{bmatrix} = \begin{bmatrix} 0,1a^{[1]} & 0 \\ 0 & 0,1a^{[1]} \end{bmatrix}$$

$$\delta^{[1]} = \frac{da^{[1]}}{dNET^{[1]}} \times (W^{[2]})^T \times \delta^{[2]}$$

$$= \begin{bmatrix} 0{,}1a^{[1]} & 0 \\ 0 & 0{,}1a^{[1]} \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times \delta^{[2]}$$

$$= \begin{bmatrix} 0{,}1a^{[1]} \\ 0{,}1a^{[1]} \end{bmatrix} \times \delta^{[2]}$$

$x = 0{,}8 \rightarrow \delta^{[2]} = -3{,}172$

$x = 1 \rightarrow \delta^{[2]} = -2{,}623$

$x = 1{,}2 \rightarrow \delta^{[2]} = -1{,}217$

$$\sum \delta^{[2]} = -7{,}012$$

$$b^{[2]new} = 1 - (-7{,}012 \times 0{,}1) = 1{,}7012$$

$$x = 0{,}8 \rightarrow \delta^{[1]} = \begin{bmatrix} 0{,}1 \times 1{,}197 \\ 0{,}1 \times 1{,}197 \end{bmatrix} \times -3{,}172 = \begin{bmatrix} -0{,}380 \\ -0{,}380 \end{bmatrix}$$

$$x = 1 \rightarrow \delta^{[1]} = \begin{bmatrix} 0{,}1 \times 1{,}221 \\ 0{,}1 \times 1{,}221 \end{bmatrix} \times -2623 = \begin{bmatrix} -0{,}320 \\ -0{,}320 \end{bmatrix}$$

$$x = 1{,}2 \rightarrow \delta^{[1]} = \begin{bmatrix} 0{,}1 \times 1{,}246 \\ 0{,}1 \times 1{,}246 \end{bmatrix} \times -1{,}217 = \begin{bmatrix} -0{,}152 \\ -0{,}152 \end{bmatrix}$$

$$\sum \delta^{[2]} = \begin{bmatrix} -0{,}852 \\ -0{,}852 \end{bmatrix} \quad b^{[1]new} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}852 \\ -0{,}852 \end{bmatrix}$$

$$= \begin{bmatrix} 1{,}085 \\ 1{,}085 \end{bmatrix}$$

$x = 0,8: \; \frac{dl}{dw^{[2]}} = \delta^{[2]} \times a^{[1]T} = -3,172 \times \begin{bmatrix} 1,197 & 1,197 \end{bmatrix}$
$= \begin{bmatrix} -3,797 & -3,797 \end{bmatrix}$

$x = 1: -2,623 \times \begin{bmatrix} 1,221 & 1,221 \end{bmatrix} = \begin{bmatrix} -3,203 & -3,203 \end{bmatrix}$

$x = 1,2: -1,217 \times \begin{bmatrix} 1,246 & 1,246 \end{bmatrix} = \begin{bmatrix} -1,516 & -1516 \end{bmatrix}$

$\sum \frac{dl}{dw^{[2]}} = \begin{bmatrix} -8,516 & -8,516 \end{bmatrix} \quad w_2^{[2]new} =$
$= \begin{bmatrix} 1 & 1 \end{bmatrix} - 0,1 \begin{bmatrix} -8,516 & -8,516 \end{bmatrix}$
$= \begin{bmatrix} 1,8516 & 1,8516 \end{bmatrix}$

$x = 0,8: \; \delta^{[1]} \times a^{[0]T} = \begin{bmatrix} -0,380 \\ -0,380 \end{bmatrix} \times 0,8 = \begin{bmatrix} -0,304 \\ -0,304 \end{bmatrix}$

$x = 1: \; \delta \begin{bmatrix} -0,320 \\ -0,320 \end{bmatrix} \times 1 = \begin{bmatrix} -0,320 \\ -0,320 \end{bmatrix}$

$x = 1,2: \; \begin{bmatrix} -0,152 \\ -0,152 \end{bmatrix} \times 1,2 = \begin{bmatrix} -0,182 \\ -0,182 \end{bmatrix}$

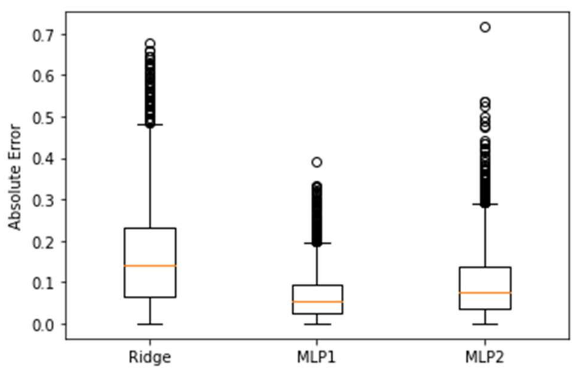
$\sum \frac{dl}{dw^{[1]}} = \begin{bmatrix} -0,806 \\ -0,806 \end{bmatrix} \quad w^{[1]new} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0,1 \begin{bmatrix} -0,806 \\ -0,806 \end{bmatrix} =$
$= \begin{bmatrix} 1,0806 \\ 1,0806 \end{bmatrix}$

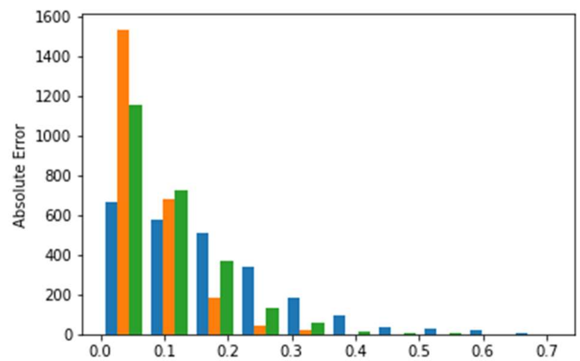
Aprendizagem 2022/23
Homework III – Group 102

II. Programming and critical analysis

1. MEA ridge: 0.162829976437694
   MEA 1: 0.06804140737968428
   MEA 2: 0.0978071820387748

2.





3. MLP1 converged in 452 iterations.
   MLP2 converged in 77 iterations.

4. Enabling the early stop flag highly impacts the number of iterations, from 77 to 402, since enabling the flag, 10% of the samples are not studied in order to test them after learning the rest 90%. This takes longer (±330 iterations more) to get all the test set correct, temporarily reducing the accuracy in the main test set while also highly reducing any overfitting problem that might appear permanently.

III. APPENDIX

```
import numpy as np, pandas as pd, matplotlib.pyplot as plt
from scipy.io.arff import loadarff
from sklearn.model_selection import train_test_split

from sklearn.metrics import mean_absolute_error
from sklearn.linear_model import Ridge
from sklearn.neural_network import MLPRegressor
from matplotlib.pyplot import hist
from matplotlib.pyplot import boxplot
#import dataset
data = loadarff('drive/MyDrive/ML/kin8nm.arff')
df = pd.DataFrame(data[0])
X = df.drop('y', axis=1)
y = df['y']
#split train/test set
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.7, random_state = 0)
# Ridge regression
rng = Ridge(alpha = 0.1)
rng.fit(X_train, y_train)
```

```
y_pred_rng = rng.predict(X_test)
print("MEA ridge:", mean_absolute_error(y_test, y_pred_rng))
#MLP 1
mlp1 = MLPRegressor(random_state=0, activation = "tanh", max_iter=500, hidden_layer_sizes = (10,10),
 early_stopping=True)
mlp1.fit(X_train.values, y_train.values)
y_pred1 = mlp1.predict(X_test.values)
print("MEA 1:", mean_absolute_error(y_test, y_pred1))
print("MLP1 converged in {} iterations.".format(mlp1.n_iter_))
#MPL 2
mlp2 = MLPRegressor(random_state=0, activation = "tanh", max_iter=500, hidden_layer_sizes = (10,10),
 early_stopping=False)
mlp2.fit(X_train.values, y_train.values)
y_pred2 = mlp2.predict(X_test.values)
print("MEA 2:", mean_absolute_error(y_test, y_pred2))
print("MLP2 converged in {} iterations.".format(mlp2.n_iter_))
#boxplot
plt.boxplot([abs(y_test - y_test_pred) for y_test_pred in [y_pred_rng, y_pred1, y_pred2]])
plt.xticks([1,2,3], ["Ridge", "MLP1", "MLP2"])
plt.ylabel("Absolute Error")
plt.show()
#histogram
plt.hist([abs(y_test - y_test_pred) for y_test_pred in [y_pred_rng, y_pred1, y_pred2]])
#plt.xticks([1,2,3], ["Ridge", "MLP1", "MLP2"])
plt.ylabel("Absolute Error")
plt.show()
```

END