Aprendizagem 2022/23

Homework III – Group 084

I. Pen-and-paper

1)

$$① \; E(\omega) = \frac{1}{2} \sum_{i=1}^{n} (z_i - \omega^T \cdot x_i)^2 + \frac{\lambda}{2} \|\omega\|^2$$

Gradiente = 0 :

$$\nabla E(\omega) = \nabla \left(\frac{1}{2} \cdot (z - X \cdot \omega)^T (Z - X \cdot \omega) + \frac{\lambda}{2} \omega^T \omega \right) = 0 \Leftrightarrow$$

$$\Leftrightarrow -2X^T Z + 2X^T \cdot X \cdot \omega + 2\lambda \cdot \omega = 0 \Leftrightarrow$$

$$\Leftrightarrow X^T Z = (X^T \cdot X + \lambda \cdot I) \cdot \omega \Leftrightarrow$$

$$\Leftrightarrow \boxed{(X^T \cdot X + \lambda \cdot I)^{-1} \cdot X^T \cdot Z = \omega}$$

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

$$X^T \cdot X = \begin{bmatrix} 5 & 6 & 7.6 & 10.08 \\ 6 & 7.6 & 10.08 & 13.88 \\ 7.6 & 10.08 & 13.88 & 19.68 \\ 10.08 & 13.88 & 19.68 & 28.56 \end{bmatrix} \quad X^T \cdot X + \lambda \cdot I = \begin{bmatrix} 7 & 6 & 7.6 & 10.08 \\ 6 & 9.6 & 10.08 & 13.88 \\ 7.6 & 10.08 & 15.88 & 19.68 \\ 10.08 & 13.88 & 19.68 & 30.55 \end{bmatrix}$$

$$(X^T \cdot X + \lambda \cdot I)^{-1} = \begin{bmatrix} 0.34 & -0.12 & -0.07 & -0.01 \\ -0.12 & 0.39 & -0.10 & -0.07 \\ -0.07 & -0.10 & 0.37 & -0.17 \\ -0.01 & -0.07 & -0.17 & 0.18 \end{bmatrix} \quad Z = \begin{bmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{bmatrix} \quad X^T Z = \begin{bmatrix} 79 \\ 88.6 \\ 105.96 \\ 134.39 \end{bmatrix}$$

$$(X^T \cdot X + \lambda \cdot I)^{-1} \cdot X^T \cdot Z = \boxed{\omega = \begin{bmatrix} 7.47 \\ 5.07 \\ 1.97 \\ -0.82 \end{bmatrix}}$$

$$\hat{z}(x, \omega) = \sum_{j=0}^{3} \omega_j \, \phi_j(x) = 7.47 + 5.07x + 1.97x^2 - 0.82x^3$$

2)

(2)

$$RMSE(\hat{z}, z) = \sqrt{\frac{1}{5} \sum_{i=1}^{5} (z_i - \hat{z}_i)^2}$$

$$\hat{z}_1 = 7.47 + 5.07 \cdot 0.8 + 1.97 \cdot 0.8^2 - 0.82 \cdot 0.8^3 = 12.37$$

$$\hat{z}_2 = 7.47 + 5.07 \cdot 1 + 1.97 \cdot 1^2 - 0.82 \cdot 1^3 = 13.69$$

$$\hat{z}_3 = 7.47 + 5.07 \cdot 1.2 + 1.97 \cdot 1.2^2 - 0.82 \cdot 1.2^3 = 14.97$$

$$\hat{z}_4 = 7.47 + 5.07 \cdot 1.4 + 1.97 \cdot 1.4^2 - 0.82 \cdot 1.4^3 = 16.18$$

$$\hat{z}_5 = 7.47 + 5.07 \cdot 1.6 + 1.97 \cdot 1.6^2 - 0.82 \cdot 1.6^3 = 17.27$$

$$RMSE(\hat{z}, z) = \sqrt{\frac{1}{5} \cdot \left((24 - 12.37)^2 + (20 - 13.69)^2 + (10 - 14.97)^2 + (13 - 16.18)^2 + (12 - 17.27)^2 \right)} = 6.89$$

3)

(3) $w^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$; $w^{[2]} = \begin{bmatrix} 1 & 1 \end{bmatrix}$; $b^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$; $b^{[2]} = 1$

$$z^{[1]} = w^{[1]} x + b^{[1]} \quad ; \quad x^{[1]} = f(z^{[1]}) = e^{0.1z^{[1]}}$$

$$z^{[2]} = w^{[2]} x^{[1]} + b^{[2]} \quad ; \quad x^{[2]} = f(z^{[2]}) = e^{0.1z^{[2]}} \quad , \quad E = \frac{1}{2}(x^{[2]} - t)^2$$

$$\frac{\partial E}{\partial x^{[2]}} = x^{[2]} - t \; ; \; \frac{\partial x^{[i]}}{\partial z^{[i]}} = 0.1 e^{0.1z^{[i]}} \; ; \; \frac{\partial z^{[i]}}{\partial w^{[i]}} = x^{[i-1]} \; ; \; \frac{\partial z^{[i]}}{\partial b^{[i]}} = 1 \; ; \; \frac{\partial z^{[i]}}{\partial x^{[i-1]}} = w^{[i]}$$

$$\delta^{[2]} = \frac{\partial E}{\partial x^{[2]}} \circ \frac{\partial x^{[2]}}{\partial z^{[2]}} = (x^{[2]} - t) \circ 0.1 e^{0.1z^{[2]}}$$

$$\delta^{[1]} = \left(\frac{\partial z^{[2]}}{\partial x^{[1]}}\right)^T \cdot \delta^{[2]} \circ \frac{\partial x^{[1]}}{\partial z^{[1]}} = w^{[2]T} \cdot \delta^{[2]} \circ 0.1 e^{0.1z^{[1]}}$$

$$\frac{\partial E}{\partial w^{[i]}} = \delta^{[i]} \cdot \frac{\partial z^{[i]}}{\partial w^{[i]}} = \delta^{[i]} \cdot (x^{[i-1]})^T$$

$$\frac{\partial E}{\partial b^{[i]}} = \delta^{[i]} \cdot \frac{\partial z^{[i]}}{\partial b^{[i]}} = \delta^{[i]}$$

$$w^{[i]} = w^{[i]} - \eta \cdot \frac{\partial E}{\partial w^{[i]}} \; ; \; b^{[i]} = b^{[i]} - \eta \cdot \frac{\partial E}{\partial b^{[i]}}$$

$x^{[0]} = 0.8$: —————— // ——————

$$z^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot 0.8 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.8 \\ 1.8 \end{bmatrix} \; ; \; x^{[1]} = \begin{bmatrix} 1.1972 \\ 1.1972 \end{bmatrix}$$

$$z^{[2]} = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1.1972 \\ 1.1972 \end{bmatrix} + 1 = 3.3944 \; ; \; x^{[2]} = 1.4042$$

$$\delta^{[2]} = (x^{[2]} - t) \circ 0.1 e^{0.1z^{[2]}} = (1.4042 - 24) \cdot 0.1 \cdot 1.4042 = -3.1728$$

$$\delta^{[1]} = w^{[2]T} \cdot \delta^{[2]} \circ 0.1 e^{0.1z^{[1]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot (-3.1728) \circ 0.1 \begin{bmatrix} 1.1972 \\ 1.1972 \end{bmatrix} = \begin{bmatrix} -0.3798 \\ -0.3798 \end{bmatrix}$$

$x^{[0]} = 1$: —— // ——

$$z^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot 1 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \; ; \; x^{[1]} = \begin{bmatrix} 1.2214 \\ 1.2214 \end{bmatrix}$$

$$z^{[2]} = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1.2214 \\ 1.2214 \end{bmatrix} + 1 = 3.4428 \; ; \; x^{[2]} = 1.4110$$

$$\delta^{[2]} = (x^{[2]} - t) \circ 0.1 e^{0.1 z^{[2]}} = (1.4110 - 20) \circ 0.1 \cdot 1.4110 = -2.6229$$

$$\delta^{[1]} = w^{[2]^T} \cdot \delta^{[2]} \circ 0.1 e^{0.1 z^{[1]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot (-2.6229) \circ 0.1 \begin{bmatrix} 1.2214 \\ 1.2214 \end{bmatrix} = \begin{bmatrix} -0.3206 \\ -0.3206 \end{bmatrix}$$

$x^{[0]} = 1.2$: —— // ——

$$z^{[1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot 1.2 + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.2 \\ 2.2 \end{bmatrix} \; ; \; x^{[1]} = \begin{bmatrix} 1.2461 \\ 1.2461 \end{bmatrix}$$

$$z^{[2]} = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1.2461 \\ 1.2461 \end{bmatrix} + 1 = 3.4922 \; ; \; x^{[2]} = 1.4180$$

$$\delta^{[2]} = (x^{[2]} - t) \circ 0.1 e^{0.1 z^{[2]}} = (1.4180 - 10) \circ 0.1 \cdot 1.4180 = -1.2169$$

$$\delta^{[1]} = w^{[2]^T} \cdot \delta^{[2]} \circ 0.1 e^{0.1 z^{[1]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot (-1.2169) \circ 0.1 \begin{bmatrix} 1.2461 \\ 1.2461 \end{bmatrix} = \begin{bmatrix} -0.1516 \\ -0.1516 \end{bmatrix}$$

—— // ——

$x^{[0]} = 0.8$: $$\frac{\partial E}{\partial w^{[2]}} = \delta^{[2]} (x^{[1]})^T = -3.1728 \cdot \begin{bmatrix} 1.1972 & 1.1972 \end{bmatrix} = \begin{bmatrix} -3.7985 & -3.7985 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}} = \delta^{[1]} (x^{[0]})^T = \begin{bmatrix} -0.3798 \\ -0.3798 \end{bmatrix} \cdot 0.8 = \begin{bmatrix} -0.3039 \\ -0.3039 \end{bmatrix}$$

$x^{[0]} = 1$: $$\frac{\partial E}{\partial w^{[2]}} = \delta^{[2]} (x^{[1]})^T = -2.6229 \cdot \begin{bmatrix} 1.2214 & 1.2214 \end{bmatrix} = \begin{bmatrix} -3.2036 & -3.2036 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}} = \delta^{[1]} (x^{[0]})^T = \begin{bmatrix} -0.3206 \\ -0.3206 \end{bmatrix} \cdot 1 = \begin{bmatrix} -0.3206 \\ -0.3206 \end{bmatrix}$$

$x^{[0]} = 1.2$: $$\frac{\partial E}{\partial w^{[2]}} = \delta^{[2]} (x^{[1]})^T = -1.2169 \cdot \begin{bmatrix} 1.2461 & 1.2461 \end{bmatrix} = \begin{bmatrix} -1.5164 & -1.5164 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}} = \delta^{[1]} (x^{[0]})^T = \begin{bmatrix} -0.1516 \\ -0.1516 \end{bmatrix} \cdot 1.2 = \begin{bmatrix} -0.1820 \\ -0.1820 \end{bmatrix}$$

$$x^{(0)} = 0.8: \quad \frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} = -3.1728$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta^{[1]} = \begin{bmatrix} -0.3798 \\ -0.3798 \end{bmatrix}$$

$$x^{(0)} = 1: \quad \frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} = -2.6229$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta^{[1]} = \begin{bmatrix} -0.3206 \\ -0.3206 \end{bmatrix}$$

$$x^{(0)} = 1.2: \quad \frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} = -1.2169$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta^{[1]} = \begin{bmatrix} -0.1516 \\ -0.1516 \end{bmatrix}$$

4

$$w^{[2]} = w^{[2]} - \eta \left(\frac{\partial E}{\partial w^{[2]}} x_1^{(0)} + \frac{\partial E}{\partial w^{[2]}} x_2^{(0)} + \frac{\partial E}{\partial w^{[2]}} x_3^{(0)} \right) =$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} - 0.1 \left(0.8 \cdot \begin{bmatrix} -3.7985 & -3.7985 \end{bmatrix} + 1 \cdot \begin{bmatrix} -3.2036 & -3.2036 \end{bmatrix} + 1.2 \begin{bmatrix} -1.5164 & -1.5164 \end{bmatrix} \right) =$$

$$= \begin{bmatrix} 1.8518 & 1.8518 \end{bmatrix}$$

$$w^{[1]} = w^{[1]} - \eta \left(\frac{\partial E}{\partial w^{[1]}} x_1^{(0)} + \frac{\partial E}{\partial w^{[1]}} x_2^{(0)} + \frac{\partial E}{\partial w^{[1]}} x_3^{(0)} \right) =$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.1 \left(0.8 \begin{bmatrix} -0.3039 \\ -0.3039 \end{bmatrix} + 1 \begin{bmatrix} -0.3206 \\ -0.3206 \end{bmatrix} + 1.2 \begin{bmatrix} -0.1820 \\ -0.1820 \end{bmatrix} \right) = \begin{bmatrix} 1.0782 \\ 1.0782 \end{bmatrix}$$

$$b^{[2]} = b^{[2]} - \eta \left(\frac{\partial E}{\partial b^{[2]}} x_1^{(0)} + \frac{\partial E}{\partial b^{[2]}} x_2^{(0)} + \frac{\partial E}{\partial b^{[2]}} x_3^{(0)} \right) =$$

$$= 1 - 0.1 \left(0.8 \cdot (-3.1728) + 1 \cdot (-2.6229) + 1.2 \cdot (-1.2169) \right) = 1.6621$$

$$b^{[1]} = b^{[1]} - \eta \left(\frac{\partial E}{\partial b^{[1]}} x_1^{(0)} + \frac{\partial E}{\partial b^{[1]}} x_2^{(0)} + \frac{\partial E}{\partial b^{[1]}} x_3^{(0)} \right) =$$

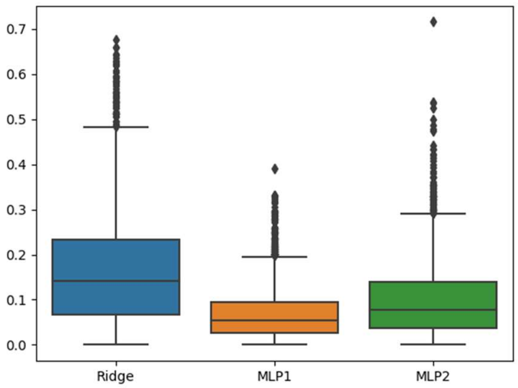
$$= 1 - 0.1 \left(0.8 \begin{bmatrix} -0.3798 \\ -0.3798 \end{bmatrix} + 1 \begin{bmatrix} -0.3206 \\ -0.3206 \end{bmatrix} + 1.2 \begin{bmatrix} -0.1516 \\ -0.1516 \end{bmatrix} \right) = \begin{bmatrix} 1.0806 \\ 1.0806 \end{bmatrix}$$

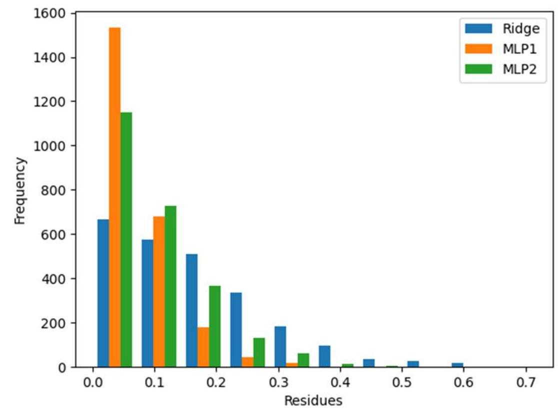
II. Programming and critical analysis

4)

MAE for Ridge Regression: 0.163;
MAE for MLP with early stopping: 0.068;
MAE for MLP without early stopping: 0.098.

5)





6)

MLP1 iterations: 452;
MLP2 iterations: 77.

7)

O MLP1 é parametrizado com early stopping, enquanto o MLP2 não. O early stopping tem como função validar o modelo, pondo de parte 10% do training data e terminando a execução quando o resultado de validação for inferior a um determinado limiar pré-definido. Logo podemos concluir que o facto do MLP1 ser parametrizado com early stopping faz com que este execute menos iterações.

III. APPENDIX

```
#######         Importing required libraries          #######
from scipy.io.arff import loadarff
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import Ridge
from sklearn.neural_network import MLPRegressor
from sklearn import metrics
import warnings

def warn(*args, **kwargs): pass
warnings.warn = warn


#######            Reading the ARFF file              #######
data = loadarff('kin8nm.arff')
df = pd.DataFrame(data[0])


#######      Creating the training-testing split      #######
X, y = df.drop('y', axis=1), df['y']
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7,
random_state=0)


#######     Creating and asserting the regressors     #######

# Creating the regressors
ridge = Ridge(alpha=0.1)
mlp1 = MLPRegressor(hidden_layer_sizes=(10,10), max_iter=500, early_stopping=True,
random_state=0, activation='tanh')
mlp2 = MLPRegressor(hidden_layer_sizes=(10,10), max_iter=500, early_stopping=False,
random_state=0, activation='tanh')

# Training the regressors
ridge.fit(X_train, y_train)
mlp1.fit(X_train, y_train)
mlp2.fit(X_train, y_train)

# Predicting the values
ridge_pred = ridge.predict(X_test)
mlp1_pred = mlp1.predict(X_test)
mlp2_pred = mlp2.predict(X_test)


#######     Calculating the MAE for each regressor    #######
ridge_mae = metrics.mean_absolute_error(y_test, ridge_pred)
mlp1_mae = metrics.mean_absolute_error(y_test, mlp1_pred)
mlp2_mae = metrics.mean_absolute_error(y_test, mlp2_pred)
```

```
#######                     Ex 4                     #######
print("MAE for Ridge Regression: ", round(ridge_mae, 3))
print("MAE for MLP with early stopping: ", round(mlp1_mae, 3))
print("MAE for MLP without early stopping: ", round(mlp2_mae, 3))


#######                     Ex 5                     #######
residues = [abs(y_test - ridge_pred), abs(y_test - mlp1_pred), abs(y_test -
mlp2_pred)]

# Boxplot
sns.boxplot(residues)
plt.xticks([0, 1, 2], ['Ridge', 'MLP1', 'MLP2'])
plt.show()

# Histogram
plt.hist(residues, bins=10, label=['Ridge', 'MLP1', 'MLP2'])
plt.legend()
plt.xlabel('Residues')
plt.ylabel('Frequency')
plt.show()


#######                     Ex 6                     #######
print("MLP1 iterations: ", mlp1.n_iter_)
print("MLP2 iterations: ", mlp2.n_iter_)
```

END