Aprendizagem 2022/23

Homework III – Group 010

I. Pen-and-paper

1) $\lambda = 2$

$$X = \varphi(x) = \begin{bmatrix} 1 & 0.8 & 0.64 & 0.512 \\ 1 & 1 & 1 & 1 \\ 1 & 1.2 & 1.44 & 1.728 \\ 1 & 1.4 & 1.96 & 2.744 \\ 1 & 1.6 & 2.56 & 4.096 \end{bmatrix}$$

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

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.8 & 1 & 1.2 & 1.4 & 1.6 \\ 0.64 & 1 & 1.44 & 1.96 & 2.56 \\ 0.512 & 1 & 1.728 & 2.744 & 4.096 \end{bmatrix} \begin{bmatrix} 1 & 0.8 & 0.64 & 0.512 \\ 1 & 1 & 1 & 1 \\ 1 & 1.2 & 1.44 & 1.728 \\ 1 & 1.4 & 1.96 & 2.744 \\ 1 & 1.6 & 2.56 & 4.096 \end{bmatrix} + 2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)^{-1} *$$

$$* \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.8 & 1 & 1.2 & 1.4 & 1.6 \\ 0.64 & 1 & 1.44 & 1.96 & 2.56 \\ 0.512 & 1 & 1.728 & 2.744 & 4.096 \end{bmatrix} * \begin{bmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{bmatrix} = \begin{bmatrix} 7.045 \\ 4.641 \\ 1.967 \\ -1.301 \end{bmatrix}$$

2)

$$\hat{z}(x) = 7.045 + 4.641x + 1.967x^2 - 1.301x^3$$

$$\hat{z}_0 = 7.045 + 4.641 * 0.8 + 1.967 * 0.64 - 1.301 * 0.512 = 11.3506$$

$$\hat{z}_1 = 7.045 + 4.641 * 1 + 1.967 * 1 - 1.301 * 1 = 12.3520$$

$$\hat{z}_2 = 7.045 + 4.641 * 1.2 + 1.967 * 1.44 - 1.301 * 1.728 = 13.1986$$

$$\hat{z}_3 = 7.045 + 4.641 * 1.4 + 1.967 * 1.96 - 1.301 * 2.744 = 13.8278$$

$$\hat{z}_4 = 7.045 + 4.641 * 1.6 + 1.967 * 2.56 - 1.301 * 4.096 = 14.1772$$

$$RMSE(\hat{z}, z) = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (z_i - \hat{z}_i)^2}$$

$$= \sqrt{\frac{(11.3506-24)^2 + (12.3520-20)^2 + (13.1986-10)^2 + (13.8278-13)^2 + (14.1722-12)^2}{5}}$$

$$= 6.843$$

Aprendizagem 2022/23

Homework III – Group 010

3) $f(x) = e^{0.1x}; f'(x) = 0.1e^{0.1x}$

$w^{[1]} = \begin{bmatrix}1\\1\end{bmatrix}; \quad b^{[1]} = \begin{bmatrix}1\\1\end{bmatrix}; \quad w^{[2]} = [1 \quad 1]; \quad b^{[2]} = [1]; \quad \eta = 0.1$

Forward Propagation:

$Z_1^{[1]} = w^{[1]} * X_1^{[0]} + b^{[1]} = \begin{bmatrix}1\\1\end{bmatrix} * [0.8] + \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1.8\\1.8\end{bmatrix};$ $\quad X_1^{[1]} = f\left(Z_1^{[1]}\right) = \begin{bmatrix}1.1972\\1.1972\end{bmatrix}$

$Z_1^{[2]} = w^{[2]} * X_1^{[1]} + b^{[2]} = [1 \quad 1] * \begin{bmatrix}1.1972\\1.1972\end{bmatrix} + [1] = [3.3944];$ $\quad X_1^{[2]} = f\left(Z_1^{[2]}\right) = [1.4042]$

$Z_2^{[1]} = w^{[1]} * X_2^{[0]} + b^{[1]} = \begin{bmatrix}1\\1\end{bmatrix} * [1] + \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\2\end{bmatrix};$ $\quad X_2^{[1]} = f\left(Z_2^{[1]}\right) = \begin{bmatrix}1.2214\\1.2214\end{bmatrix}$

$Z_2^{[2]} = w^{[2]} * X_2^{[1]} + b^{[2]} = [1 \quad 1] * \begin{bmatrix}1.2214\\1.2214\end{bmatrix} + [1] = [3.4428];$ $\quad X_2^{[2]} = f\left(Z_2^{[2]}\right) = [1.4110]$

$Z_3^{[1]} = w^{[1]} * X_3^{[0]} + b^{[1]} = \begin{bmatrix}1\\1\end{bmatrix} * [1.2] + \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2.2\\2.2\end{bmatrix};$ $\quad X_3^{[1]} = f\left(Z_3^{[1]}\right) = \begin{bmatrix}1.2461\\1.2461\end{bmatrix}$

$Z_3^{[2]} = w^{[2]} * X_3^{[1]} + b^{[2]} = [1 \quad 1] * \begin{bmatrix}1.2461\\1.2461\end{bmatrix} + [1] = [3.4922];$ $\quad X_3^{[2]} = f\left(Z_3^{[2]}\right) = [1.4180]$

Back Propagation:

$$E(w) = \frac{1}{2} * \sum_{k=1}^{n} \left(X_k^{[i]} - t\right)^2$$

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial X} \circ \frac{\partial X}{\partial Z} * \left(\frac{\partial Z}{\partial w}\right)^T = \delta * \left(\frac{\partial Z}{\partial w}\right)^T$$

$$\delta_1^{[2]} = \left(X_1^{[2]} - t_1\right) \circ \left(0.1 * X_1^{[2]}\right) = ([1.4042] - [24]) \circ (0.1 * [1.4042]) = [-3.1729]$$

$$\left(\frac{\partial Z_1^{[2]}}{\partial w_1^{[2]}}\right)^T = \left(X_1^{[1]}\right)^T = [1.1972 \quad 1.1972]$$

$$\delta_2^{[2]} = \left(X_2^{[2]} - t_2\right) \circ \left(0.1 * X_2^{[2]}\right) = ([1.4110] - [20]) \circ (0.1 * [1.4110]) = [-2.6229]$$

$$\left(\frac{\partial Z_2^{[2]}}{\partial w_2^{[2]}}\right)^T = \left(X_2^{[1]}\right)^T = [1.2214 \quad 1.2214]$$

$$\delta_3^{[2]} = \left(X_3^{[2]} - t_3\right) \circ \left(0.1 * X_3^{[2]}\right) = ([1.4180] - [10]) \circ (0.1 * [1.4180]) = [-1.2169]$$

$$\left(\frac{\partial Z_3^{[2]}}{\partial w_3^{[2]}}\right)^T = \left(X_3^{[1]}\right)^T = [1.2461 \quad 1.2461]$$

$$\frac{\partial E}{\partial w^{[2]}} = \frac{\partial E}{\partial w_1^{[2]}} + \frac{\partial E}{\partial w_2^{[2]}} + \frac{\partial E}{\partial w_3^{[2]}} =$$

$$= [-3.1729] * [1.1972 \quad 1.1972] + [-2.6229] * [1.2214 \quad 1.2214] + [-1.2169] * [1.2461 \quad 1.2461] =$$

$$= [-8.5186 \quad -8.5186]$$

$$w^{[2]} = w^{[2]} - \eta \frac{\partial E}{\partial w^{[2]}} = [1 \quad 1] - 0.1 * [-8.5186 \quad -8.5186] = [1.85186 \quad 1.85186]$$

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial X} \circ \frac{\partial X}{\partial Z} * \left(\frac{\partial Z}{\partial b}\right)^T = \delta * \left(\frac{\partial Z}{\partial b}\right)^T; \left(\frac{\partial Z}{\partial b}\right)^T = 1$$

$$\frac{\partial E}{\partial b^{[2]}} = \delta_1^{[2]} + \delta_2^{[2]} + \delta_3^{[2]} = [-3.1729] + [-2.6229] + [-1.2169] = [-7.0127]$$

$$b^{[2]} = b^{[2]} - \eta \frac{\partial E}{\partial b^{[2]}} = [1] - 0.1 * [-7.0127] = [1.70127]$$

$$\frac{\partial E}{\partial w^{[1]}} = \frac{\partial E}{\partial X^{[1]}} \circ \frac{\partial X^{[1]}}{\partial Z^{[1]}} * \left(\frac{\partial Z^{[1]}}{\partial w^{[1]}}\right)^T = \delta^{[1]} * \left(\frac{\partial Z^{[1]}}{\partial w^{[1]}}\right)^T$$

$$\delta^{[1]} = \left(\left(\frac{\partial Z^{[2]}}{\partial X^{[1]}}\right)^T * \delta^{[2]}\right) \circ \frac{\partial X^{[1]}}{\partial Z^{[1]}} = \left(\left(w^{[2]}\right)^T * \delta^{[2]}\right) \circ 0.1 * X^{[1]}$$

$$\delta_1^{[1]} = \left(\left(w_1^{[2]}\right)^T * \delta_1^{[2]}\right) \circ 0.1 * X_1^{[1]} = \left(\begin{bmatrix}1\\1\end{bmatrix} * [-3.1729]\right) \circ 0.1 * \begin{bmatrix}1.1972\\1.1972\end{bmatrix} = \begin{bmatrix}-0.3799\\-0.3799\end{bmatrix}$$

$$\left(\frac{\partial Z_1^{[1]}}{\partial w_1^{[1]}}\right)^T = \left(X_1^{[0]}\right)^T = [0.8]$$

$$\delta_2^{[1]} = \left(\left(w_2^{[2]}\right)^T * \delta_2^{[2]}\right) \circ 0.1 * X_2^{[1]} = \left(\begin{bmatrix}1\\1\end{bmatrix} * [-2.6229]\right) \circ 0.1 * \begin{bmatrix}1.2214\\1.2214\end{bmatrix} = \begin{bmatrix}-0.3204\\-0.3204\end{bmatrix}$$

$$\left(\frac{\partial Z_2^{[1]}}{\partial w_2^{[1]}}\right)^T = \left(X_2^{[0]}\right)^T = [1]$$

$$\delta_3^{[1]} = \left(\left(w_3^{[2]}\right)^T * \delta_3^{[2]}\right) \circ 0.1 * X_3^{[1]} = \left(\begin{bmatrix}1\\1\end{bmatrix} * [-1.2169]\right) \circ 0.1 * \begin{bmatrix}1.2461\\1.2461\end{bmatrix} = \begin{bmatrix}-0.1516\\-0.1516\end{bmatrix}$$

$$\left(\frac{\partial Z_3^{[1]}}{\partial w_3^{[1]}}\right)^T = \left(X_3^{[0]}\right)^T = [1.2]$$

$$\frac{\partial E}{\partial w^{[1]}} = \frac{\partial E}{\partial w_1^{[1]}} + \frac{\partial E}{\partial w_2^{[1]}} + \frac{\partial E}{\partial w_3^{[1]}} =$$

$$= \begin{bmatrix}-0.3799\\-0.3799\end{bmatrix}[0.8] + \begin{bmatrix}-0.3204\\-0.3204\end{bmatrix}[1] + \begin{bmatrix}-0.1516\\-0.1516\end{bmatrix}[1.2] = \begin{bmatrix}-0.8062\\-0.8062\end{bmatrix}$$

$$w^{[1]} = w^{[1]} - \eta \frac{\partial E}{\partial w^{[1]}} = \begin{bmatrix}1\\1\end{bmatrix} - 0.1 * \begin{bmatrix}-0.8062\\-0.8062\end{bmatrix} = \begin{bmatrix}1.08062\\1.08062\end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta_1^{[1]} + \delta_2^{[1]} + \delta_3^{[1]} = \begin{bmatrix}-0.3799\\-0.3799\end{bmatrix} + \begin{bmatrix}-0.3204\\-0.3204\end{bmatrix} + \begin{bmatrix}-0.1516\\-0.1516\end{bmatrix} = \begin{bmatrix}-0.8519\\-0.8519\end{bmatrix}$$

$$b^{[1]} = b^{[1]} - \eta \frac{\partial E}{\partial b^{[1]}} = \begin{bmatrix}1\\1\end{bmatrix} - 0.1 * \begin{bmatrix}-0.8519\\-0.8519\end{bmatrix} = \begin{bmatrix}1.08519\\1.08519\end{bmatrix}$$

II. Programming and critical analysis

4) MAE(Ridge): 0.162829976437694

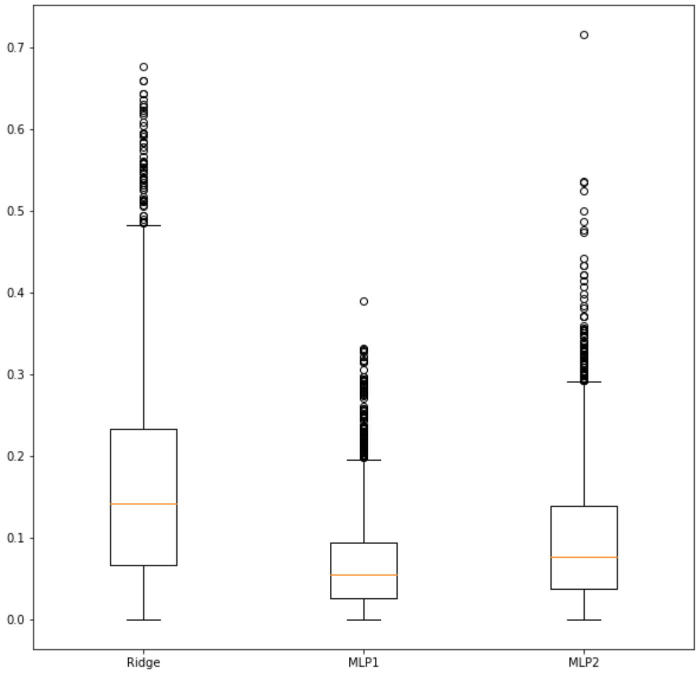
MAE(MPL1): 0.0680414073796843

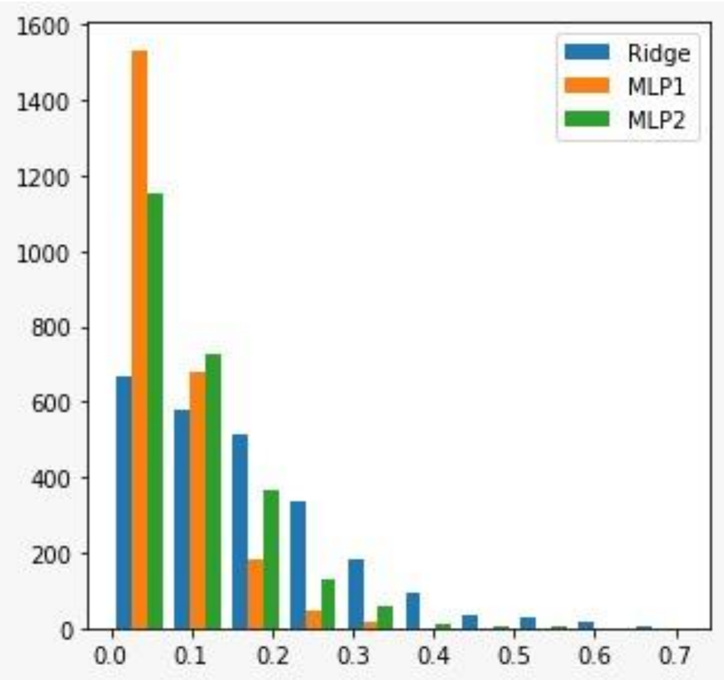
MAE(MPL2): 0.0978071820387748

Aprendizagem 2022/23

Homework III – Group 010

5)





6) MLP1 iterations: 452

MLP2 iterations: 77

7) The regressor with early stopping will separate 10% of the input data for validation. This validation data will decide when the convergence algorithm should stop (in this case, it stops when there are *10* iterations with a validation score less than *1e-4*, default values of *n_iter_no_change* and *tol* attributes respectively). The regressor without early stopping only uses the training loss on the entire input data for the stopping criterion. For this reason, early stopping regressor may have more iterations than the non-early stopping regressor, as in this case. Additionally, we can notice that MLP1's MAE is smaller than MLP2's MAE, and this means that the goal of decrease overfitting with early stopping was achieved.

III. APPENDIX

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

from sklearn.linear_model import Ridge
from sklearn.neural_network import MLPRegressor

data = loadarff('kin8nm.arff')
df = pd.DataFrame(data[0])

target = df['y']

df = df.drop('y', axis=1)

x_train, x_test, y_train, y_test = train_test_split(df.values, target, test_size = 0.3,\
random_state=0)

rr = Ridge(alpha=0.1)

mlp1 = MLPRegressor(hidden_layer_sizes=(10,10), activation='tanh', random_state=0,\
       max_iter=500, early_stopping=True)
mlp2 = MLPRegressor(hidden_layer_sizes=(10,10), activation='tanh', random_state=0,\
       max_iter=500, early_stopping=False)

#4
rr.fit(x_train, y_train)
mlp1.fit(x_train, y_train)
mlp2.fit(x_train, y_train)

y_pred_rr = rr.predict(x_test)
y_pred_mlp1 = mlp1.predict(x_test)
y_pred_mlp2 = mlp2.predict(x_test)

print("MAE(Ridge): ", metrics.mean_absolute_error(y_test, y_pred_rr))
print("MAE(MPL1): ", metrics.mean_absolute_error(y_test, y_pred_mlp1))
print("MAE(MPL2): ", metrics.mean_absolute_error(y_test, y_pred_mlp2))

#5.1
plt.figure(figsize=(5, 5))
boxplot = plt.boxplot([abs(y_test - y_pred_rr), abs(y_test - y_pred_mlp1), \
       abs(y_test - y_pred_mlp2)], labels=['Ridge', 'MLP1', 'MLP2'])
```

```
#5.2
plt.figure(figsize=(5, 5))

hist = plt.hist([abs(y_test - y_pred_rr), abs(y_test - y_pred_mlp1), \
       abs(y_test - y_pred_mlp2)])
plt.legend(labels=['Ridge', 'MLP1', 'MLP2'])

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

END