Aprendizagem 2021/22

Homework III – Group 105

I. Pen-and-paper

1)

$$x^{(1)} = [0.8] \quad x^{(2)} = [1] \quad x^{(3)} = [1.2] \quad x^{(4)} = [1.4] \quad x^{(5)} = [1.6]$$

$$t^{(1)} = [24] \quad t^{(2)} = [20] \quad t^{(3)} = [10] \quad t^{(4)} = [13] \quad t^{(5)} = [12]$$

$$\Phi_j(x) = x^j \quad , \quad j = 0, 1, 2, 3$$

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

$$W = \left(\Phi^T \cdot \Phi + \lambda \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)^{-1} \cdot \Phi^T \cdot \begin{bmatrix} 24 \\ 20 \\ 10 \\ 13 \\ 12 \end{bmatrix}$$

$$\omega = \left(\begin{bmatrix} 5 & 6 & 7.6 & 10.08 \\ 6 & 7.6 & 10.08 & 13.8784 \\ 7.6 & 10.08 & 13.8784 & 19.68 \\ 10.08 & 13.8784 & 19.68 & 28.55488 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \right) \cdot \begin{bmatrix} 79 \\ 88.6 \\ 105.96 \\ 134.392 \end{bmatrix}$$

$$\omega = \begin{bmatrix} 7.0450759 \\ 4.64092765 \\ 1.96734046 \\ -1.30088142 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 105

2)

$$\hat{z}(x_0, w) = 7.0450759 \times 1 + 4.64092765 \times 0.8 + 1.96734046 \times 0.8^2 - 1.30088142 \times 0.8^3$$

$$= 11.3508646$$

$$\hat{z}(x_1, w) = 7.0450759 \times 1 + 4.64092765 \times 1 + 1.96734046 \times 1 - 1.30088142 \times 1$$

$$= 12.35246259$$

$$\hat{z}(x_2, w) = 7.0450759 \times 1 + 4.64092765 \times 1.2 + 1.96734046 \times 1.2^2 - 1.30088142 \times 1.2^3$$

$$= 13.19923625$$

$$\hat{z}(x_3, w) = 7.0450759 \times 1 + 4.64092765 \times 1.4 + 1.96734046 \times 1.4^2 - 1.30088142 \times 1.4^3$$

$$= 13.8287433$$

$$\hat{z}(x_4, w) = 7.0450759 \times 1 + 4.64092765 \times 1.6 + 1.96734046 \times 1.6^2 - 1.30088142 \times 1.6^3$$

$$= 14.17854143$$

$$W^T \cdot \Phi^T = \left[11.35086463,\ 12.35246259,\ 13.19923625,\ 13.8287433,\ 14.17854143 \right]$$

$$RMSE = \sqrt{\frac{(24 - 11.35086463)^2 + (20 - 12.35246259)^2 + (10 - 13.19923625)^2 + (13 - 13.8287433)^2 + (12 - 14.17854143)^2}{5}}$$

$$= 6.84329489134$$

3)

$$x^{(0)} = \begin{bmatrix} 0.8 & 1 & 1.2 \end{bmatrix} \quad t = \begin{bmatrix} 24 & 20 & 10 \end{bmatrix} \quad f(x) = e^{0.1x}$$

$$b^{(1)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad W^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \eta = 0.1$$

$$b^{(2)} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \quad W^{(2)} = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

foward propagation

$$z^{(n)} = W^{(n)} x^{(n-1)} + b^{(n)}$$

$$z^{(1)} = W^{(1)} x^{(0)} + b^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} . \begin{bmatrix} 0.8 & 1 & 1.2 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.8 & 1 & 1.2 \\ 0.8 & 1 & 1.2 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1.8 & 2 & 2.2 \\ 1.8 & 2 & 2.2 \end{bmatrix}$$

$$x^{(1)} = f(x = z^{(1)}) = \begin{bmatrix} e^{0.18} & e^{0.2} & e^{0.22} \\ e^{0.18} & e^{0.2} & e^{0.22} \end{bmatrix}$$

$$= \begin{bmatrix} 1.1972 & 1.2214 & 1.2461 \\ 1.1972 & 1.2214 & 1.2461 \end{bmatrix}$$

$$z^{(2)} = W^{(2)} . x^{(1)} + b^{(2)} = \begin{bmatrix} 1 & 1 \end{bmatrix} . \begin{bmatrix} 1.1972 & 1.2214 & 1.2461 \\ 1.1972 & 1.2214 & 1.2461 \end{bmatrix} +$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2.3944 & 2.4428 & 2.4922 \end{bmatrix} =$$

$$= \begin{bmatrix} 3.3944 & 3.4428 & 3.4922 \end{bmatrix}$$

$$x^{(2)} = f(u = z^{(2)}) = \begin{bmatrix} e^{0.33944} & e^{0.34428} & e^{0.34922} \end{bmatrix} = \begin{bmatrix} 1.4042 & 1.4110 & 1.4180 \end{bmatrix}$$

back propagation

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

$$\frac{\partial E(x^{(2)}, t)}{\partial x^{(2)}} = \frac{\partial \frac{1}{2}\sum (x^{(2)} - t)^2}{\partial x^{(2)}} = \sum (x^{(2)} - t)^2 =$$

$$= \begin{bmatrix} 1.4042 - 24 & 1.4110 - 20 & 1.4180 - 10 \end{bmatrix} = \begin{bmatrix} -22.5958 & -18.589 & -8.582 \end{bmatrix}$$

$$\frac{\partial x^{(i)}}{\partial z^{(i)}} = \frac{\partial f(x = z^{(i)})}{\partial z^{(i)}} = 0.1 f(z^{(i)})$$

$$\frac{\partial z^{(i)}}{\partial x^{(i-1)}} = \omega^{(i)}$$

$$\delta^2 = \frac{\partial E}{\partial x^{(2)}} \circ \frac{\partial x^{(2)}}{\partial t^{(2)}} = \sum (x^{(2)} - t) \circ 0.1 f(z^{(2)}) =$$

$$= \begin{bmatrix} -22.5958 & -18.589 & -8.582 \end{bmatrix} \circ \begin{bmatrix} 0.14042 & 0.14110 & 0.14180 \end{bmatrix}$$

$$= \begin{bmatrix} -3.1729 & -2.6230 & -1.2170 \end{bmatrix}$$

$$\delta^{(1)} = \left(\frac{\partial z^{(2)}}{\partial x^{(1)}}\right)^T \cdot \delta^{(2)} \circ \frac{\partial x^{(1)}}{\partial z^{(1)}} = (\omega^{(2)})^T \cdot \delta^2 \circ 0.1\, f(z^{(1)})$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3.1729 & -2.6230 & -1.2170 \end{bmatrix} \circ \begin{bmatrix} 0.11972 & 0.12214 & 0.12461 \\ 0.11972 & 0.12214 & 0.12461 \end{bmatrix}$$

$$= \begin{bmatrix} -3.1729 & -2.6230 & -1.2170 \\ -3.1729 & -2.6230 & -1.2170 \end{bmatrix} \circ \begin{bmatrix} 0.11972 & 0.12214 & 0.12461 \\ 0.11972 & 0.12214 & 0.12461 \end{bmatrix}$$

$$= \begin{bmatrix} -0.3799 & -0.3204 & -0.1517 \\ -0.3799 & -0.3204 & -0.1517 \end{bmatrix}$$

upgrades

$$\omega^{(i)}_{new} = \omega^{(i)} - \eta \frac{\partial E}{\partial \omega^{(i)}} = \omega^{(i)} - 0.1 . \delta^{(i)} . (x^{(i-1)})^T$$

$$\omega^{(1)}_{new} = \omega^{(1)} - 0.1 . \delta^{(1)} . (x^{(0)})^T =$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -0.03799 & -0.03204 & -0.01517 \\ -0.03799 & -0.03204 & -0.01517 \end{bmatrix} . \begin{bmatrix} 0.8 \\ 1 \\ 1.2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -0.080636 \\ -0.080636 \end{bmatrix}$$

$$= \begin{bmatrix} 1.080636 \\ 1.080636 \end{bmatrix}$$

$$\omega^{(2)}_{new} = \omega^{(2)} - 0.1\,\delta^{(2)}\left(x^{(1)}\right)^T =$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} - \begin{bmatrix} -0.31729 & -0.26230 & -0.12170 \end{bmatrix} \cdot \begin{bmatrix} 1.1972 & 1.1972 \\ 1.2214 & 1.2214 \\ 1.2461 & 1.2461 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} - \begin{bmatrix} -0.851883 & -0.851883 \end{bmatrix} = \begin{bmatrix} 1.851883 & 1.851883 \end{bmatrix}$$

$$b^{(i)}_{new} = b^{(i)} - \eta\frac{\partial E}{\partial b^{(i)}} = b^{(i)} - 0.1\sum_{i}^{3}\delta^{(i)}$$

$$b^{(1)}_{new} = b^{(1)} - 0.1\sum_{i=1}^{3}\delta^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -0.0852 \\ -0.0852 \end{bmatrix}$$

$$= \begin{bmatrix} 1.0852 \\ 1.0852 \end{bmatrix}$$

$$b^{(2)}_{new} = b^{(2)} - 0.1\sum_{i=1}^{3}\delta^{(2)} = [1] - [0.70129] = [1.70129]$$

Aprendizagem 2021/22

Homework III – Group 105

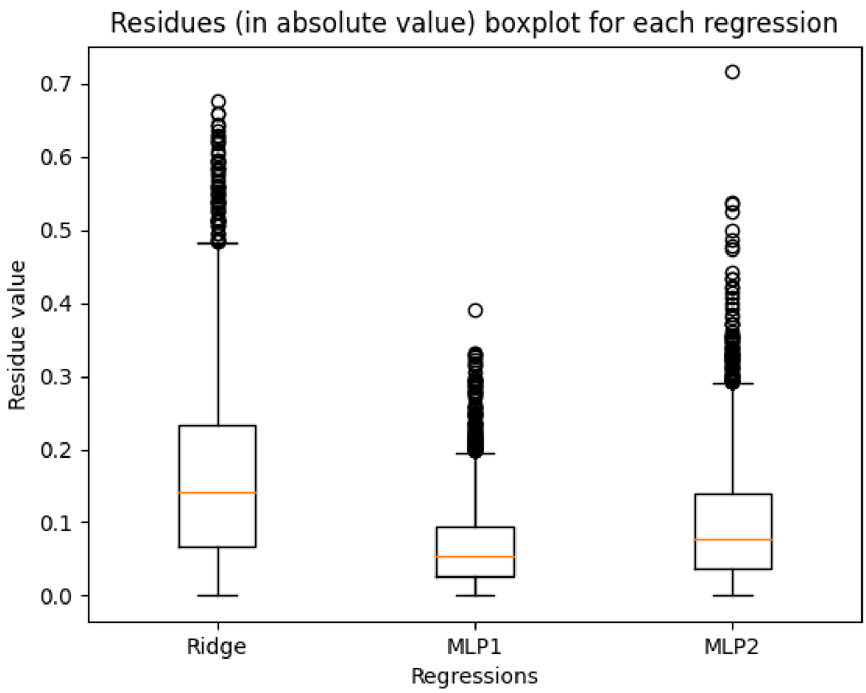
II. Programming and critical analysis

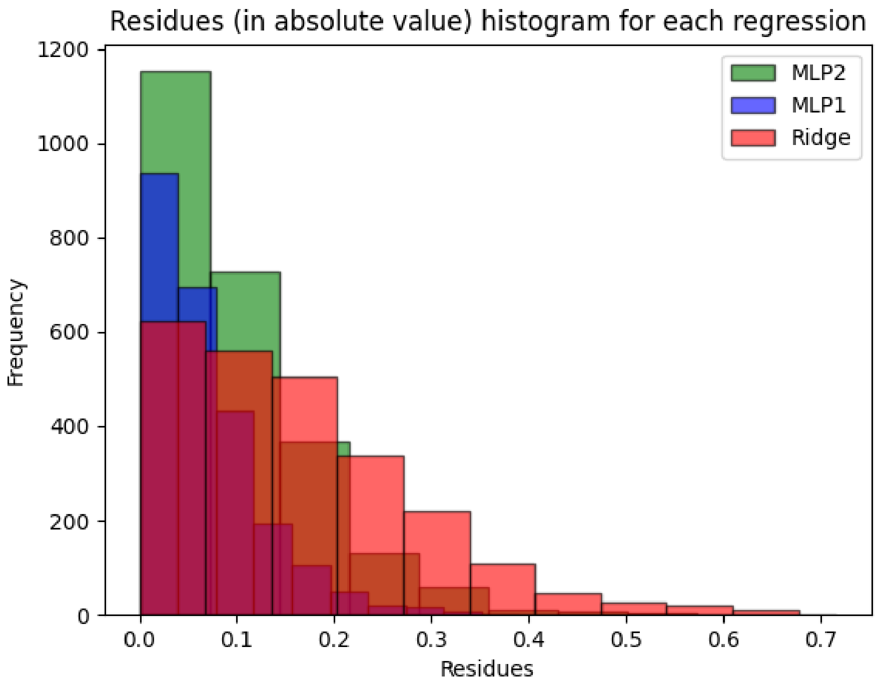
4) Ridge MAE: 0.162829976437694

MLP1 MAE: 0.0680414073796843

MLP2 MAE: 0.0978071820387748

5)





6) MLP_1: 452

MLP_2: 77

7) RMSE (Ridge): 0.2036848188708104

R^2 (Ridge): 0.3988587401698873

RMSE (MLP1): 0.08909520405604758

R^2 (MLP1): 0.8849814552439308

RMSE (MLP2): 0.127294627272514

R2 (MLP2): 0.7652101262308029

Observando as diferenças de valores de RMSE, R^2 e MAE, podemos observar que ambos os MLP tiveram um desempenho bastante superior à regressão Ridge, dentro dos quais podemos verificar um leve aumento de desempenho do MLP1 (com early stopping) em relação ao MLP2 (sem early stopping). Analisando o número de iterações, podemos ver que o MLP1 precisou de mais do quíntuplo das iterações do MLP2, facto que pode ser justificado devido à presença/ausência de early stopping. Ao utilizar early stopping, a regressão irá calcular a loss através dos dados de validação, de forma a adaptar a MLP a dados novos o que, obviamente, irá levar a um maior número de iterações. Assim, podemos verificar um trade-off claro entre ter ou não early stopping, ao utilizar esta estratégia, teremos uma performance superior e diminuímos a chance de over-fitting (pois expomos a MLP a novos dados), a custo de um maior número de iterações e, consequentemente, necessitaremos de mais tempo para dar fit da MLP. Sem early stopping teremos resultados inversos, ou seja, menor performance, mas menos iterações e, portanto, menos tempo de fit.

III. APPENDIX

```
import warnings

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.model_selection import StratifiedKFold, cross_val_score, train_test_split,
learning_curve
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from scipy import stats
from scipy.io.arff import loadarff
from sklearn.naive_bayes import GaussianNB
import seaborn as sns
from sklearn.preprocessing import normalize
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.neural_network import MLPRegressor

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

# Reading the ARFF file
data = loadarff('kin8nm.arff')
df = pd.DataFrame(data[0])
X, y = df[list(df.columns[:-1])], df[["y"]]

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

```
rr = Ridge(alpha=0.1)
mlpr_1 = MLPRegressor(hidden_layer_sizes=(10,10), activation="tanh", random_state=0, max_iter=500,
early_stopping=True)
mlpr_2 = MLPRegressor(hidden_layer_sizes=(10,10), activation="tanh", random_state=0, max_iter=500)

rr.fit(X_train, y_train)
mlpr_1.fit(X_train, y_train)
mlpr_2.fit(X_train, y_train)

y_test_rr = rr.predict(X_test)
y_test_mlpr_1 = mlpr_1.predict(X_test)
y_test_mlpr_2 = mlpr_2.predict(X_test)

mae_rr = mean_absolute_error(y_test, y_test_rr)
mae_mlpr_1 = mean_absolute_error(y_test, y_test_mlpr_1)
mae_mlpr_2 = mean_absolute_error(y_test, y_test_mlpr_2)

rr_residuals = []
mlpr_1_residuals = []
mlpr_2_residuals = []

size = len(y_test)
for i in range(size):
    rr_residuals.append(abs(y_test.iloc[i]["y"] - y_test_rr[i][0]))
    mlpr_1_residuals.append(abs(y_test.iloc[i]["y"] - y_test_mlpr_1[i]))
    mlpr_2_residuals.append(abs(y_test.iloc[i]["y"] - y_test_mlpr_2[i]))

n_iteration_mlpr_1 = mlpr_1.niter
n_iteration_mlpr_2 = mlpr_2.niter

print("MLP1 iterations:", n_iteration_mlpr_1)
print("MLP2 iterations:", n_iteration_mlpr_2)

print("RMSE (Ridge):",np.sqrt(mean_squared_error(y_test,y_test_rr)))
print("R2 (Ridge):",r2_score(y_test, y_test_rr))

print("RMSE (MLP1):",np.sqrt(mean_squared_error(y_test,y_test_mlpr_1)))
print("R2 (MLP1):",r2_score(y_test, y_test_mlpr_1))

print("RMSE (MLP2):",np.sqrt(mean_squared_error(y_test,y_test_mlpr_2)))
print("R2 (MLP2):",r2_score(y_test, y_test_mlpr_2))

plt.boxplot([rr_residuals, mlpr_1_residuals, mlpr_2_residuals], labels=["Ridge", "MLP1", "MLP2"])
plt.title("Residues (in absolute value) boxplot for each regression")
plt.xlabel("Regressions")
plt.ylabel("Residue value")
plt.show()

plt.hist(mlpr_2_residuals, edgecolor="black", linewidth=1, color="green", alpha=0.60, label="MLP2",
align="mid")
plt.hist(mlpr_1_residuals, edgecolor="black", linewidth=1, color="blue", alpha=0.60, label="MLP1",
align="mid")
plt.hist(rr_residuals, edgecolor="black", linewidth=1, color="red", alpha=0.60, label="Ridge",
align="mid")

plt.title("Residues (in absolute value) histogram for each regression")
plt.xlabel("Residues")
plt.ylabel("Frequency")
plt.legend()
plt.show()
print("Ridge MAE:", mae_rr, "\nMLP1 MAE:", mae_mlpr_1, "\nMLP2 MAE:", mae_mlpr_2)
```

END