Aprendizagem 2022/23

Homework III – Group 009

I. Pen-and-paper

1) A closed form solution da regressão de Ridge com λ=2, sobre o espaço de dados transformado é dada por:

$$w=(\phi^T\phi+\lambda I)^{-1}\phi^T z$$

Onde:

$$\phi=\begin{bmatrix}1 & 0.8 & 0.8^2 & 0.8^3\\ 1 & 1 & 1^2 & 1^3\\ 1 & 1.2 & 1.2^2 & 1.2^3\\ 1 & 1.4 & 1.4^2 & 1.4^3\\ 1 & 1.6 & 1.6^2 & 1.6^3\end{bmatrix}\quad z=\begin{bmatrix}24\\20\\10\\13\\12\end{bmatrix}$$

Substituindo temos:

$$w=\left(\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 0.8 & 1 & 1.2 & 1.4 & 1.6\\ 0.64 & 1 & 1.44 & 1.96 & 2.56\\ 0.512 & 1 & 1.728 & 2.744 & 4.096\end{bmatrix}\begin{bmatrix}1 & 0.8 & 0.64 & 0.512\\ 1 & 1 & 1 & 1\\ 1 & 1.2 & 1.44 & 1.728\\ 1 & 1.4 & 1.96 & 2.744\\ 1 & 1.6 & 2.56 & 4.096\end{bmatrix}+2\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\right)^{-1}\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 0.8 & 1 & 1.2 & 1.4 & 1.6\\ 0.64 & 1 & 1.44 & 1.96 & 2.56\\ 0.512 & 1 & 1.728 & 2.744 & 4.096\end{bmatrix}\begin{bmatrix}24\\20\\10\\13\\12\end{bmatrix}$$

$$w=\left(\begin{bmatrix}5 & 6 & 7.6 & 10.08\\ 6 & 7.6 & 10.08 & 13.8784\\ 7.6 & 10.08 & 13.8784 & 19.68\\ 10.08 & 13.8784 & 19.68 & 28.5549\end{bmatrix}+\begin{bmatrix}2&0&0&0\\0&2&0&0\\0&0&2&0\\0&0&0&2\end{bmatrix}\right)^{-1}\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 0.8 & 1 & 1.2 & 1.4 & 1.6\\ 0.64 & 1 & 1.44 & 1.96 & 2.56\\ 0.512 & 1 & 1.728 & 2.744 & 4.096\end{bmatrix}\begin{bmatrix}24\\20\\10\\13\\12\end{bmatrix}$$

$$w=\left(\begin{bmatrix}7 & 6 & 7.6 & 10.08\\ 6 & 9.6 & 10.08 & 13.8784\\ 7.6 & 10.08 & 15.8784 & 19.68\\ 10.08 & 13.8784 & 19.68 & 30.5549\end{bmatrix}\right)^{-1}\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 0.8 & 1 & 1.2 & 1.4 & 1.6\\ 0.64 & 1 & 1.44 & 1.96 & 2.56\\ 0.512 & 1 & 1.728 & 2.744 & 4.096\end{bmatrix}\begin{bmatrix}24\\20\\10\\13\\12\end{bmatrix}$$

$$w=\begin{bmatrix}0.3417 & -0.1214 & -0.0749 & -0.0093\\ -0.1214 & 0.3892 & -0.0967 & -0.0745\\ -0.0749 & -0.0967 & 0.3726 & -0.1714\\ -0.0093 & -0.0745 & -0.1714 & 0.1701\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 0.8 & 1 & 1.2 & 1.4 & 1.6\\ 0.64 & 1 & 1.44 & 1.96 & 2.56\\ 0.512 & 1 & 1.728 & 2.744 & 4.096\end{bmatrix}\begin{bmatrix}24\\20\\10\\13\\12\end{bmatrix}$$

$$w=\begin{bmatrix}0.1918 & 0.1360 & 0.0720 & -0.0007 & -0.0825\\ 0.0899 & 0.0966 & 0.0777 & 0.0297 & -0.0512\\ -0.0015 & 0.0296 & 0.0495 & 0.0498 & 0.0224\\ -0.0864 & -0.0751 & -0.0344 & 0.0445 & 0.1701\end{bmatrix}\begin{bmatrix}24\\20\\10\\13\\12\end{bmatrix}$$

$$w=\begin{bmatrix}7.0451\\4.6409\\1.9673\\-1.3009\end{bmatrix}$$

Portanto:

$$\hat{z}(x)=7.0451+4.6409\,x+1.9673\,x^2-1.3009\,x^3$$

2) O RMSE é dado por:

$$RMSE=\sqrt{\frac{\sum(\hat{z}-z)^2}{n}}$$

Tendo em conta os seguintes valores de ẑ:

$$\hat{z}(0.8)=7.0451+4.6409(0.8)+1.9673(0.8)^2-1.3009(0.8)^3=11.3509$$

$$\hat{z}(1)=7.0451+4.6409(1)+1.9673(1)^2-1.3009(1)^3=12.3525$$

$$\hat{z}(1.2)=7.0451+4.6409(1.2)+1.9673(1.2)^2-1.3009(1.2)^3=13.1992$$

$$\hat{z}(1.4)=7.0451+4.6409(1.4)+1.9673(1.4)^2-1.3009(1.4)^3=13.8287$$

$$\hat{z}(1.6)=7.0451+4.6409(1.6)+1.9673(1.6)^2-1.3009(1.6)^3=14.1785$$

Ficamos com:

$$RMSE=\sqrt{\frac{(11.3509-24)^2+(12.3525-20)^2+(13.1992-10)^2+(13.8287-13)^2+(14.1785-12)^2}{5}}$$

$$RMSE=\sqrt{\frac{234.1534}{5}}=6.8433$$

3) Tendo:

$$f(x)=e^{0.1x},\ \eta=0.1,\ w^{[1]}=\begin{bmatrix}1\\1\end{bmatrix},\ b^{[1]}=\begin{bmatrix}1&1&1\\1&1&1\end{bmatrix},\ w^{[2]}=\begin{bmatrix}1&1\end{bmatrix},\ b^{[2]}=\begin{bmatrix}1&1&1\end{bmatrix},\ X=\begin{bmatrix}0.8&1&1.2\end{bmatrix}$$

Começamos por fazer forward propagation:

$$z^{[1]}=w^{[1]}X+b^{[1]}\Leftrightarrow z^{[1]}=\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}0.8&1&1.2\end{bmatrix}+\begin{bmatrix}1&1&1\\1&1&1\end{bmatrix}=\begin{bmatrix}0.8&1&1.2\\0.8&1&1.2\end{bmatrix}+\begin{bmatrix}1&1&1\\1&1&1\end{bmatrix}=\begin{bmatrix}1.8&2&2.2\\1.8&2&2.2\end{bmatrix}$$

$$v^{[1]}=f\left(\begin{bmatrix}1.8&2&2.2\\1.8&2&2.2\end{bmatrix}\right)=\begin{bmatrix}e^{0.18}&e^{0.2}&e^{0.22}\\e^{0.18}&e^{0.2}&e^{0.22}\end{bmatrix}$$

$$z^{[2]}=w^{[2]}v^{[1]}+b^{[2]}\Leftrightarrow z^{[2]}=\begin{bmatrix}1&1\end{bmatrix}\begin{bmatrix}e^{0.18}&e^{0.2}&e^{0.22}\\e^{0.18}&e^{0.2}&e^{0.22}\end{bmatrix}+\begin{bmatrix}1&1&1\end{bmatrix}\Leftrightarrow z^{[2]}=\begin{bmatrix}2e^{0.18}&2e^{0.2}&2e^{0.22}\end{bmatrix}+\begin{bmatrix}1&1&1\end{bmatrix}$$

$$z^{[2]}=\begin{bmatrix}1+2e^{0.18} & 1+2e^{0.2} & 1+2e^{0.22}\end{bmatrix}$$

$$v^{[2]}=f\left(\begin{bmatrix}1+2e^{0.18} & 1+2e^{0.2} & 1+2e^{0.22}\end{bmatrix}\right)=\begin{bmatrix}e^{0.1(1+2e^{0.18})} & e^{0.1(1+2e^{0.2})} & e^{0.1(1+2e^{0.22})}\end{bmatrix}=\begin{bmatrix}1.4042 & 1.411 & 1.418\end{bmatrix}$$

Tendo em conta o seguinte erro:

$$E=\frac{1}{2}(z-v^{[2]})^2$$

E sabendo que a atualização dos pesos é dada por:

$$w_{new}^{[i]}=w_{old}^{[i]}-\eta\frac{\partial E}{\partial w^{[i]}}$$

E a atualização dos bias é dada por:

$$b_{new}^{[i]}=b_{old}^{[i]}-\eta\frac{\partial E}{\partial b^{[i]}}$$

Começamos por computar as seguintes derivadas:

$$\frac{\partial E}{\partial w^{[2]}}=\frac{\partial E}{\partial v^{[2]}}\frac{\partial v^{[2]}}{\partial z^{[2]}}\frac{\partial z^{[2]}}{\partial w^{[2]}}$$

$$\frac{\partial E}{\partial w^{[1]}}=\delta_2\frac{\partial z^{[2]}}{\partial v^{[1]}}\frac{\partial v^{[1]}}{\partial z^{[1]}}\frac{\partial z^{[1]}}{\partial w^{[1]}},\quad \delta_2=\frac{\partial E}{\partial v^{[2]}}\frac{\partial v^{[2]}}{\partial z^{[2]}}$$

$$\frac{\partial E}{\partial b^{[2]}}=\frac{\partial E}{\partial v^{[2]}}\frac{\partial v^{[2]}}{\partial z^{[2]}}\frac{\partial z^{[2]}}{\partial b^{[2]}}$$

$$\frac{\partial E}{\partial b^{[1]}}=\delta_2\frac{\partial z^{[2]}}{\partial v^{[1]}}\frac{\partial v^{[1]}}{\partial z^{[1]}}\frac{\partial z^{[1]}}{\partial b^{[1]}},\quad \delta_2=\frac{\partial E}{\partial v^{[2]}}\frac{\partial v^{[2]}}{\partial z^{[2]}}$$

Onde:

$$\frac{\partial E}{\partial v^{[2]}}=-(z-v^{[2]})$$

$$\frac{\partial v^{[2]}}{\partial z^{[2]}}=0.1\,f(z^{[2]})=0.1\,v^{[2]}$$

$$\frac{\partial z^{[2]}}{\partial w^{[2]}}=v^{[1]}$$

$$\frac{\partial z^{[2]}}{\partial v^{[1]}}=w^{[2]}$$

Aprendizagem 2022/23

Homework III – Group 009

$$\frac{\partial v^{[1]}}{\partial z^{[1]}}=0.1\,f(z^{[1]})=0.1\,v^{[1]}$$

$$\frac{\partial z^{[1]}}{\partial w^{[1]}}=X$$

$$\frac{\partial z^{[2]}}{\partial b^{[2]}}=1$$

$$\frac{\partial z^{[1]}}{\partial b^{[1]}}=1$$

<u>Portanto para $w^{[2]}$ temos:</u>

$$\frac{\partial E}{\partial w^{[2]}}=-(z-v^{[2]})\circ(0.1\,v^{[2]})v^{[1]T}=-\left(\begin{bmatrix}24 & 20 & 10\end{bmatrix}-\begin{bmatrix}1.4042 & 1.411 & 1.418\end{bmatrix}\right)\circ\left(0.1\begin{bmatrix}1.4042 & 1.411 & 1.418\end{bmatrix}\right)\begin{bmatrix}e^{0.18} & e^{0.18}\\ e^{0.2} & e^{0.2}\\ e^{0.22} & e^{0.22}\end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[2]}}=-\left(\begin{bmatrix}22.5958 & 18.5890 & 8.582\end{bmatrix}\right)\circ\left(\begin{bmatrix}0.14042 & 0.1411 & 0.1418\end{bmatrix}\right)\begin{bmatrix}e^{0.18} & e^{0.18}\\ e^{0.2} & e^{0.2}\\ e^{0.22} & e^{0.22}\end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[2]}}=\begin{bmatrix}-3.1728 & -2.6229 & -1.2169\end{bmatrix}\begin{bmatrix}e^{0.18} & e^{0.18}\\ e^{0.2} & e^{0.2}\\ e^{0.22} & e^{0.22}\end{bmatrix}=\begin{bmatrix}-8.5185 & -8.5185\end{bmatrix}$$

$$w^{[2]}_{new}=w^{[2]}_{old}-0.1\,\frac{\partial E}{\partial w^{[2]}}=\begin{bmatrix}1 & 1\end{bmatrix}-0.1\begin{bmatrix}-8.5185 & -8.5185\end{bmatrix}=\begin{bmatrix}1.85185 & 1.85185\end{bmatrix}$$

<u>Para $b^{[2]}$ temos:</u>

$$\frac{\partial E}{\partial b^{[2]}}=\delta_2\,x\,1=\delta_2=\begin{bmatrix}-3.1728 & -2.6229 & -1.2169\end{bmatrix}$$

$$b^{[2]}_{new}=b^{[2]}_{old}-0.1\,\frac{\partial E}{\partial b^{[2]}}=\begin{bmatrix}1 & 1 & 1\end{bmatrix}-0.1\begin{bmatrix}-3.1728 & -2.6229 & -1.2169\end{bmatrix}=\begin{bmatrix}1.31728 & 1.26229 & 1.12169\end{bmatrix}$$

<u>Para $w^{[1]}$ temos:</u>

$$\frac{\partial E}{\partial w^{[1]}}=(w^{[2]T}\,\delta_2)\circ(0.1\,v^{[1]})\,X^T=\left(\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}-3.1728 & -2.6229 & -1.2169\end{bmatrix}\right)\circ\left(0.1\begin{bmatrix}e^{0.18} & e^{0.2} & e^{0.22}\\ e^{0.18} & e^{0.2} & e^{0.22}\end{bmatrix}\right)\begin{bmatrix}0.8\\1\\1.2\end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}}=\left(\begin{bmatrix}-3.1728 & -2.6229 & -1.2169\\ -3.1728 & -2.6229 & -1.2169\end{bmatrix}\right)\circ\left(\begin{bmatrix}0.1\,e^{0.18} & 0.1\,e^{0.2} & 0.1\,e^{0.22}\\ 0.1\,e^{0.18} & 0.1\,e^{0.2} & 0.1\,e^{0.22}\end{bmatrix}\right)\begin{bmatrix}0.8\\1\\1.2\end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}}=\begin{bmatrix}-0.3799 & -0.3204 & -0.1516\\ -0.3799 & -0.3204 & -0.1516\end{bmatrix}\begin{bmatrix}0.8\\ 1\\ 1.2\end{bmatrix}=\begin{bmatrix}-0.8062\\ -0.8062\end{bmatrix}$$

$$w_{new}^{[1]}=w_{old}^{[1]}-0.1\,\frac{\partial E}{\partial w^{[1]}}=\begin{bmatrix}1\\ 1\end{bmatrix}-0.1\begin{bmatrix}-0.8062\\ -0.8062\end{bmatrix}=\begin{bmatrix}1.08062\\ 1.08062\end{bmatrix}$$

<u>**Para $b^{[1]}$ temos:**</u>

$$\frac{\partial E}{\partial b^{[1]}}=\delta_1 x 1=\delta_1=\begin{bmatrix}-0.3799 & -0.3204 & -0.1516\\ -0.3799 & -0.3204 & -0.1516\end{bmatrix}$$

$$b_{new}^{[1]}=b_{old}^{[1]}-0.1\,\frac{\partial E}{\partial b^{[1]}}=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}-0.1\begin{bmatrix}-0.3799 & -0.3204 & -0.1516\\ -0.3799 & -0.3204 & -0.1516\end{bmatrix}=\begin{bmatrix}1.03799 & 1.03204 & 1.01516\\ 1.03799 & 1.03204 & 1.01516\end{bmatrix}$$

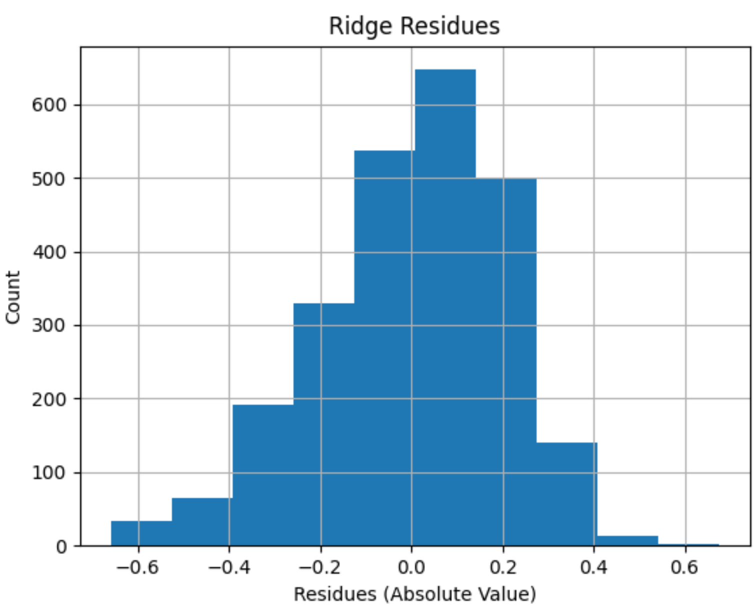
II. Programming and critical analysis

4)

$$MAE_{Ridge}=0.1628$$
$$MAE_{MLP_1}=0.0680$$
$$MAE_{MLP_2}=0.0978$$

5) <u>**Histogramas:**</u>

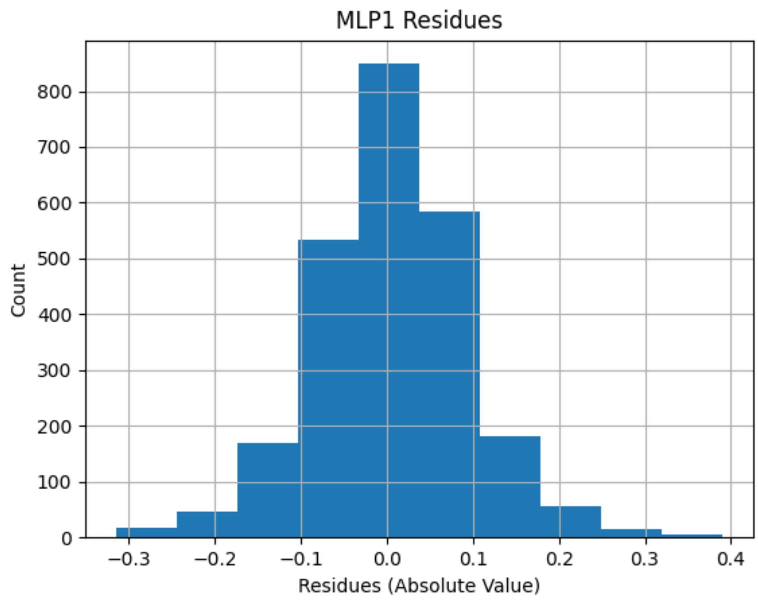
<u>**Ridge:**</u>



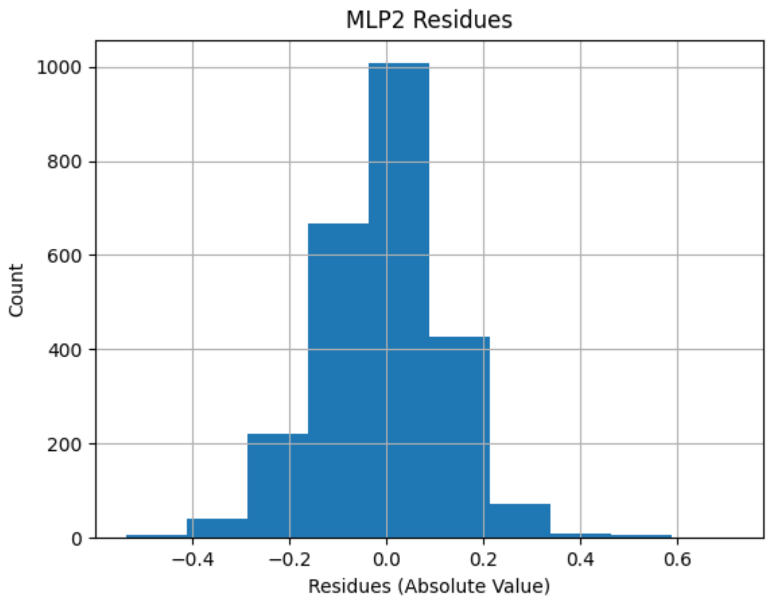
Aprendizagem 2022/23

Homework III – Group 009

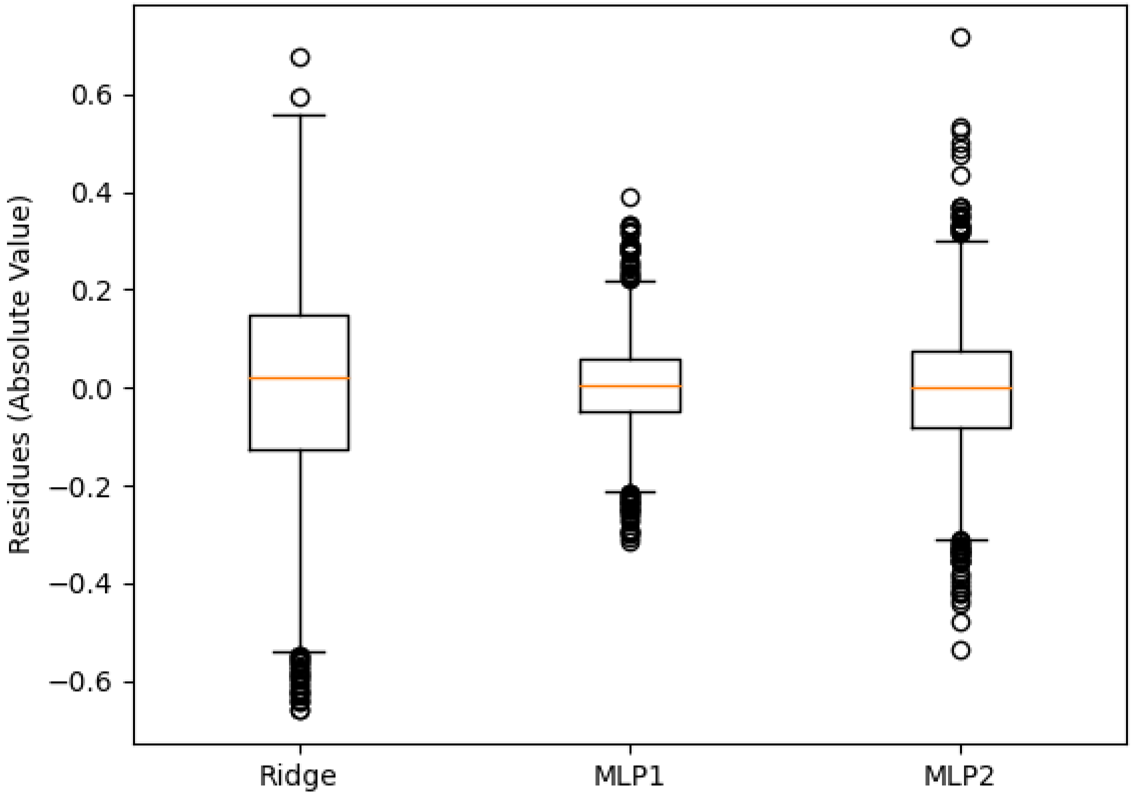
MLP$_1$:



MLP$_2$:



Boxplot:



6)

$$Iterations_{MLP_1} = 452$$

$$Iterations_{MLP_2} = 77$$

7) A principal diferença entre o MLP_1 e o MLP_2 reside no critério de convergência, onde o MLP_1 utiliza Early Stopping e o MLP_2 não, o que pode levar a um número de iterações diferente entre os dois regressores.

No MLP_1 paramos de treinar o modelo quando o desempenho do modelo deixa de melhorar no conjunto de dados de validação, enquanto que no MLP_2 o critério de paragem é quando a loss function tem consecutivas iterações em que não melhora em pelo menos uma certa tolerância definida(neste caso n_iter_no_change=10 e tol=10-4) .

Tendo em conta que o critério de convergência do Early Stopping utiliza um conjunto de dados de validação, a probabilidade de termos problemas de underfitting e de overfitting vai ser menor no MLP1. Portanto tendo em conta que $MAE_{MLP1} < MAE_{MLP2}$, $Iterations_{MLP1} > Iterations_{MLP2}$ e que MLP_1 utiliza Early Stopping, uma razão possível para as diferenças entre os MLPs pode ser estarmos a obter um modelo underfitting com o MLP_2.

III. APPENDIX

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

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)

ridge = Ridge(alpha=0.1)
ridge.fit(X_train.values,y_train)
y_test_pred_ridge = ridge.predict(X_test.values)
mae_ridge = mean_absolute_error(y_test, y_test_pred_ridge)
residues_ridge = y_test-y_test_pred_ridge

mlp1 = MLPRegressor(hidden_layer_sizes=(10,10,), activation='tanh', max_iter=500,
early_stopping=True, random_state=0)
mlp1.fit(X_train.values,y_train)
y_test_pred_mlp1 = mlp1.predict(X_test.values)
mae_mlp1 = mean_absolute_error(y_test, y_test_pred_mlp1)
residues_mlp1 = y_test-y_test_pred_mlp1

mlp2 = MLPRegressor(hidden_layer_sizes=(10,10,), activation='tanh', max_iter=500,
early_stopping=False, random_state=0)
mlp2.fit(X_train.values,y_train)
y_test_pred_mlp2 = mlp2.predict(X_test.values)
mae_mlp2 = mean_absolute_error(y_test, y_test_pred_mlp2)
residues_mlp2 = y_test-y_test_pred_mlp2

print("MAE Ridge", mae_ridge)
print("MAE MLP1", mae_mlp1)
print("MAE MLP2",mae_mlp2)

plt.hist(residues_ridge)
plt.xlabel('Residues (Absolute Value)')
plt.ylabel('Count')
plt.title('Ridge Residues')
plt.grid(True)
plt.show()

plt.hist(residues_mlp1)
plt.xlabel('Residues (Absolute Value)')
plt.ylabel('Count')
plt.title('MLP1 Residues')
plt.grid(True)
plt.show()

plt.hist(residues_mlp2)
plt.xlabel('Residues (Absolute Value)')
plt.ylabel('Count')
plt.title('MLP2 Residues')
plt.grid(True)
plt.show()

plt.boxplot([residues_ridge,residues_mlp1,residues_mlp2], labels=['Ridge','MLP1','MLP2'])
plt.ylabel('Residues (Absolute Value)')
```

```
plt.show()

print("MLP1 Iterations", mlp1.n_iter_)
print("MLP2 Iterations", mlp2.n_iter_)
```

END