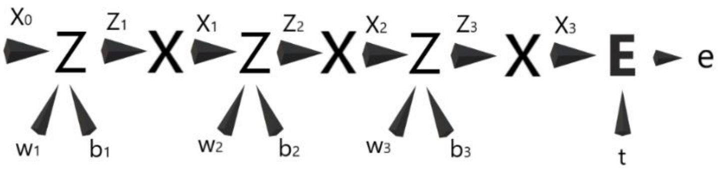
Aprendizagem 2021/22

Homework III – Group 96

I. Pen-and-paper

1)

a)



Forward: $z^l = w^l . x^{l-1} + b^l$, $x^l = \tanh(z^l)$, $e = E(x^3, T) = \frac{1}{2}||x^3 - t||_2^2$

Backward: $\frac{dE}{dx^3} = x^3 - t, \frac{dz^l}{dx^{l-1}} = w^l, \frac{dz^l}{dw^l} = x^{l-1}, \frac{dz^l}{db^l} = 1, \frac{dx^l}{dz^l} = 1 - \tanh^2((z^l)$,

$$\tanh'(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)' = \frac{(e^x + e^{-x})(e^x + e^{-x})}{(e^x + e^{-x})^2} - \frac{(e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2} = 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = 1 - \left(\frac{(e^x - e^{-x}))}{(e^x + e^{-x})}\right)^2$$

$$= 1 - \tanh^2(x)$$

$$x^0 = \begin{pmatrix}1 & 1 & 1 & 1 & 1\end{pmatrix}^\top, \quad w^1 = \begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 0 & 0 & 0 & 0 & 0\\ 1 & 1 & 1 & 1 & 1\end{pmatrix}, \; b^1 = \begin{pmatrix}1\\1\\1\end{pmatrix}, W^2 = \begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}, \; b^2 = \begin{pmatrix}1\\1\end{pmatrix}, t = \begin{pmatrix}1\\-1\end{pmatrix}$$

$$z^1 = w^1 \cdot x^0 + b^1 = \begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 0 & 0 & 0 & 0 & 0\\ 1 & 1 & 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}5\\0\\5\end{pmatrix} + \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}6\\1\\6\end{pmatrix}$$

$$W^3 = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} b^3 = \begin{pmatrix}0\\0\end{pmatrix} t = \begin{pmatrix}1\\-1\end{pmatrix},$$

$$x^1 = \tanh(z^1) = \begin{pmatrix}\tanh(6)\\ \tanh(1)\\ \tanh(6)\end{pmatrix} = \begin{pmatrix}0{,}999\\ 0{,}761\\ 0{,}999\end{pmatrix}$$

$$z^2 = w^2 \cdot x^1 + b^2 = \begin{pmatrix}3{,}761\\ 3{,}761\end{pmatrix}, x^2 = \tanh(z^2) = \begin{pmatrix}0{,}998\\ 0{,}998\end{pmatrix},$$

$$z^3 = w^3 \cdot x^2 + b^3 = \begin{pmatrix}0\\0\end{pmatrix}, x^3 = \tanh(z^3) = \begin{pmatrix}0\\0\end{pmatrix}, e = \frac{1}{2}\left\|\begin{bmatrix}0\\0\end{bmatrix} - \begin{bmatrix}1\\-1\end{bmatrix}\right\|_2^2 = 1$$

$$\partial^3 = \frac{\partial E}{\partial x^3} o \frac{\partial x^3}{\partial z^3} = (x^3 - t) o (1 - \tanh^2(z^3)) = \left(\begin{pmatrix}0\\0\end{pmatrix} - \begin{pmatrix}1\\-1\end{pmatrix}\right) o \left(\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}tanh^2(0)\\ tanh^2(0)\end{pmatrix}\right) = \begin{pmatrix}-1\\1\end{pmatrix}$$

$$\partial^2 = \left(\frac{\partial z^3}{\partial x^2}\right)^\top \cdot \partial^3 o \frac{\partial x^2}{\partial z^2} = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} \cdot \begin{pmatrix}-1\\1\end{pmatrix} o \left(\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}\tanh^2(3{,}761)\\ \tanh^2(3{,}761)\end{pmatrix}\right) = \begin{pmatrix}0\\0\end{pmatrix}$$

$$\partial^1 = \left(\frac{\partial z^2}{\partial x^1}\right)^\top \cdot \partial^2 o \frac{\partial x^1}{\partial z^1} = \begin{pmatrix}1 & 1\\ 1 & 1\\ 1 & 1\end{pmatrix} \cdot \begin{pmatrix}0\\0\end{pmatrix} o \left(\begin{pmatrix}1\\1\\1\end{pmatrix} - \begin{pmatrix}\tanh^2(6)\\ \tanh^2(1)\\ \tanh^2(6)\end{pmatrix}\right) = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

$$W^1 = W^1 - \eta \frac{\partial E}{\partial w^1} = W^1 - \eta \left(\partial^1 \cdot \left(\frac{\partial z^1}{\partial w^1} \right)^{\mathsf{T}} \right) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} - 0{,}1 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix} \right)$$

$$= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$b^1 = b^1 - \eta \frac{\partial E}{\partial b^1} = b^1 - \eta \left(\partial^1 \cdot \left(\frac{\partial z}{\partial b^1} \right)^{\mathsf{T}} \right) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 0{,}1 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$W^2 = W^2 - n \left(\partial^2 \cdot \left(\frac{\partial z^2}{\partial w^2} \right)^{\mathsf{T}} \right) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} - 0{,}1 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \cdot (0{,}999; 0{,}701; 0{,}999) \right) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$b^2 = b^2 - n \left(\partial^2 \cdot \left(\frac{\partial z^2}{\partial b^2} \right)^{\mathsf{T}} \right) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0{,}1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$w^3 = w^3 - \eta \left(\partial^3 \cdot \left(\frac{\partial z^3}{\partial \omega^3} \right)^{\mathsf{T}} \right) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} - 0{,}1 \left(\begin{pmatrix} -1 \\ 1 \end{pmatrix} \cdot (0{,}998; 0{,}998) \right) = \begin{pmatrix} 0{,}0999 & 0{,}0999 \\ -0{,}0999 & -0{,}0999 \end{pmatrix}$$

$$b^3 = b^3 - \eta \left(\partial^3 \cdot \left(\frac{\partial z^3}{\partial b^3} \right)^{\mathsf{T}} \right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0{,}1 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0{,}1 \\ -0{,}1 \end{pmatrix}$$

b) $x^3 = \text{softmax}\,(z^3), x_i = \frac{\exp(z_i)}{\sum_{k=1}^n \exp\,(z_k)}, e = E(x^3, t) = -\sum_{i=1}^n t_i \log_2 x_i^3$

$$\frac{\partial}{\partial z_j} \frac{\exp(z_i)}{\sum_k \exp\,(z_k)} = \frac{\exp\,(z_i) \sum_x \exp\,(z_k) - \exp\,(z_i) \exp\,(z_j)}{\left(\sum_x \exp(z_k)\right)^2} = \frac{\exp(z_i)}{\sum_k \exp\,(z_k)} \frac{\sum_k \exp\,(z_k) - \exp\,(z_j)}{\sum_k \exp\,(z_k)}$$

$$= \frac{\exp(z_i)}{\sum_k \exp\,(z_k)} \left(1 - \frac{\exp(z_j)}{\sum_k \exp\,(z_k)} \right) = x_i (1 - x_j), \text{se } i = j$$

$$\frac{\partial}{\partial z_j} \frac{\exp(z_i)}{\sum_k \exp\,(z_k)} = \frac{0 - \exp\,(z_j) \exp\,(z_i)}{\left(\sum_k \exp\,(z_k)\right)^2} = -\frac{\exp(z_j)}{\sum_k \exp(z_k)} \frac{\exp(z_i)}{\sum_k \exp(z_k)} = -x_j x_i, \text{se } i \neq j$$

$\frac{\partial x^3}{\partial z^3} = \begin{cases} x_i(1 - x_j), & i = j \\ -x_j x_i, & i \neq j \end{cases}, \frac{\partial E}{\partial z_i^3} = -\sum_{j=1}^n t_j \frac{1}{x_j^3} \frac{\partial x_j^3}{\partial z_i^3} = -\frac{t_i}{x_i^3} \frac{\partial x_i^3}{\partial z^3} - \sum_{j \neq 1} \frac{t_j}{x_j^3} \frac{\partial x_j^3}{\partial z_i^3} = -\frac{t_i}{x_i^3} x_i^3 (1 - x_i^3) - \sum_{j \neq 1} \frac{t_j}{x_j^3} (-x_j^3 x_i^3) = -t_i + t_i x_i^3 + \sum_{j \neq 1}^i t_j x_i^3 = -t_i + x_i^3 (t_i + \sum_{j \neq 1} t_j) = -t_i + x_i^3 (\sum_{j=1} t_j) = x_i^3 - t_i$

As restantes fórmulas continuam iguais ao ex. 1. Na *forward propagation*, só muda o x^3.

$$x^3 = \text{softmax}(z^3) = \text{softmax} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0{,}5 \\ 0{,}5 \end{pmatrix}$$

$$\partial^3 = \frac{\partial E}{\partial z^3} = x^3 - 9 = \begin{pmatrix}0{,}5\\0{,}5\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}-0{,}5\\+0{,}5\end{pmatrix}$$

$$\partial^2 = \left(\frac{\partial z^3}{\partial x^2}\right)^T \cdot \partial^3 \circ \frac{\partial x^2}{\partial z^2} = \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix} \cdot \begin{pmatrix}-0{,}5\\+0{,}5\end{pmatrix} \circ \left(\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}\tanh^2(3{,}761)\\\tanh^2(3{,}761)\end{pmatrix}\right) = \begin{pmatrix}0\\0\end{pmatrix}$$

$$\partial^1 = \left(\frac{\partial z^2}{\partial x^1}\right)^T \cdot \partial^2 \circ \frac{\partial x^1}{\partial z^1} = \begin{pmatrix}1 & 1\\1 & 1\\1 & 1\end{pmatrix} \cdot \begin{pmatrix}0\\0\end{pmatrix} \circ \left(\begin{pmatrix}1\\1\\1\end{pmatrix} - \begin{pmatrix}\tanh^2(6)\\\tanh^2(1)\\\tanh^2(6)\end{pmatrix}\right) = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

$$w^1 = w^1 - n(\partial^1 \cdot (x^0)^T) = \begin{pmatrix}1 & 1 & 1 & 1 & 1\\0 & 0 & 0 & 0 & 0\\1 & 1 & 1 & 1 & 1\end{pmatrix} - 0{,}1\left(\begin{pmatrix}0\\0\\0\end{pmatrix} \cdot \begin{pmatrix}1 & 1 & 1 & 1 & 1\end{pmatrix}\right)$$

$$= \begin{pmatrix}1 & 1 & 1 & 1 & 1\\0 & 0 & 0 & 0 & 0\\1 & 1 & 1 & 1 & 1\end{pmatrix}$$

$$b^1 = b^1 - \eta(\partial^1 \cdot 1) = \begin{pmatrix}1\\1\\1\end{pmatrix} - 0{,}1\begin{pmatrix}0\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$w^2 = w^2 - \eta(\partial^2 \cdot (x^1)^T) = \begin{pmatrix}1 & 1 & 1\\1 & 1 & 1\end{pmatrix} - 0{,}1\left(\begin{pmatrix}0\\0\end{pmatrix} \cdot (0{,}999; 0{,}761; 0{,}999)\right) = \begin{pmatrix}1 & 1 & 1\\1 & 1 & 1\end{pmatrix}$$

$$b^2 = b^2 - \eta(\partial^2 \cdot 1) = \begin{pmatrix}1\\1\end{pmatrix} - 0{,}1\begin{pmatrix}0\\0\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$$

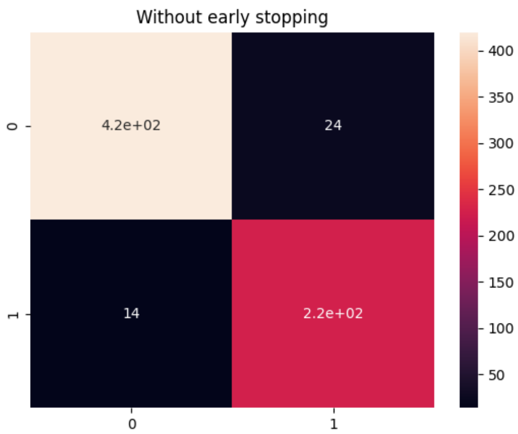
$$w^3 = w^3 - \eta(\partial^3 \cdot (x^2)^T) = \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix} - 0{,}1\left(\begin{pmatrix}-0{,}5\\+0{,}5\end{pmatrix} \cdot (0{,}998; 0{,}998)\right) = \begin{pmatrix}0{,}0499 & 0{,}0499\\-0{,}0499 & -0{,}0499\end{pmatrix}$$

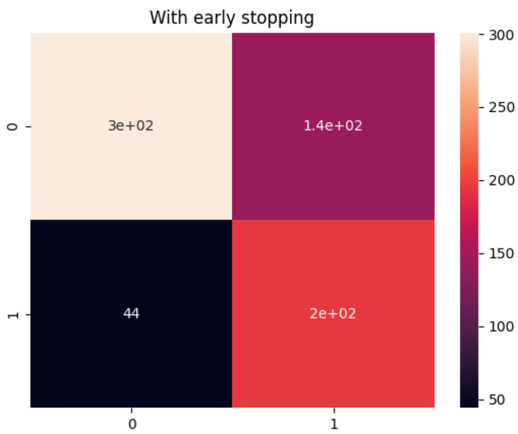
$$b^3 = b^3 - \eta(\partial^3 \cdot 1) = \begin{pmatrix}0\\0\end{pmatrix} - 0{,}1\begin{pmatrix}-0{,}5\\+0{,}5\end{pmatrix} = \begin{pmatrix}0{,}05\\-0{,}05\end{pmatrix}$$

$$e = -(1 \times \log_2 0{,}5 + 0 \times \log_2 0{,}5) = 1$$

II. Programming and critical analysis

2) Com *early stopping* obtemos resultados piores, porque, como acabamos a análise dos dados de teste mais cedo, deixamos de ter tanta especificidade, havendo mais falhas nos dados de teste. Além disso, o nosso MLP tem poucas *layers*, apenas duas, o que pode levar o modelo a não aprender os padrões apresentados nas *features* de modo a classificar bem os dados, levando assim a um desempenho pior nos dados de teste.

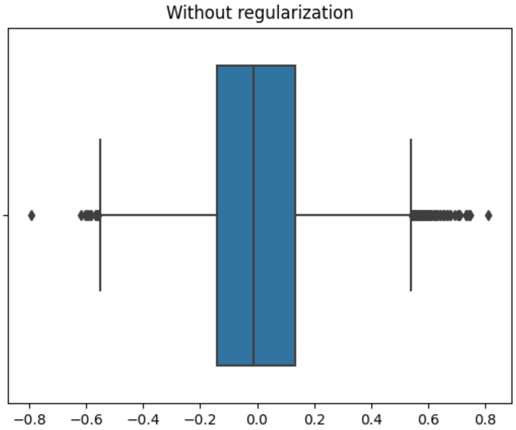


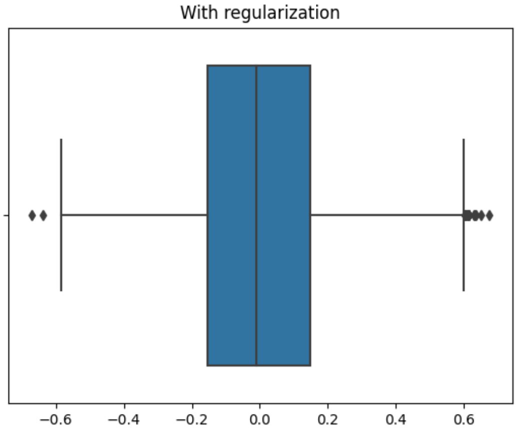


Aprendizagem 2021/22

Homework III – Group 96

3) Para reduzir o erro do MLP podemos aplicar as seguintes estratégias: maior *l2 regularization*, maior *dataset* de modo a aumentar também o *training set*, usar *early stopping* e ainda fazer mais passagens sobre os dados (aumentando o *cross-validation*).





III. APPENDIX

```
# pergunta 2
from sklearn.neural_network import MLPClassifier
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import confusion_matrix

df = pd.read_csv("test.csv", dtype={"Clump_Thickness": int,
                                    "Cell_Size_Uniformity": int,
                                    "Cell_Shape_Uniformity": int,
                                    "Marginal_Adhesion": int,
                                    "Single_Epi_Cell_Size": int,
                                    "Bare_Nuclei": float,
                                    "Bland_Chromatin": int,
                                    "Normal_Nucleoli": int,
                                    "Mitoses": int,
                                    "Class": str}, na_values=["?"])

df = df.dropna()

columns = ["Clump_Thickness", "Cell_Size_Uniformity", "Cell_Shape_Uniformity",
"Marginal_Adhesion",
           "Single_Epi_Cell_Size", "Bare_Nuclei", "Bland_Chromatin",
"Normal_Nucleoli", "Mitoses", "Class"]
features = columns[:-1]
classes = ["benign", "malign"]

Y = df["Class"]
X = df.drop(columns=["Class"])

Y = [0 if x == "benign" else 1 for x in Y]
Y = np.ravel(pd.DataFrame(Y))
```

Aprendizagem 2021/22
Homework III – Group 96

```
kf = StratifiedKFold(n_splits=5, random_state=132, shuffle=True)

# l2 regularization/alpha a default
clf = MLPClassifier(activation="relu", hidden_layer_sizes=(3, 2), random_state=1)
clf2 = MLPClassifier(activation="relu", hidden_layer_sizes=(3, 2), random_state=1,
early_stopping=True)

mat = [[0, 0], [0, 0]]
mat2 = [[0, 0], [0, 0]]
for train_index, test_index in kf.split(X, Y):
    x_train = X.iloc[train_index].loc[:, features].values
    x_test = X.iloc[test_index][features].values
    y_train = Y[train_index]
    y_test = Y[test_index]

    clf.fit(x_train, y_train)
    y_pred = clf.predict(x_test)

    clf2.fit(x_train, y_train)
    y_pred2 = clf2.predict(x_test)

    mat += confusion_matrix(y_test, y_pred)
    mat2 += confusion_matrix(y_test, y_pred2)

print(mat)
print(mat2)

sns.heatmap(mat, annot=True)
plt.title("Without early stopping")
plt.show()

sns.heatmap(mat2, annot=True)
plt.title("With early stopping")
plt.show()
```

```
# pergunta 3
import pandas as pd
from sklearn.neural_network import MLPRegressor
from sklearn.model_selection import KFold
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv("kin.csv", dtype={"theta1": float, "theta2": float, "theta3":
float, "theta4": float, "theta5": float,
                                    "theta6": float, "theta7": float, "theta8":
float, "y": float})

features = ["theta1", "theta2", "theta3", "theta4", "theta5", "theta6", "theta7",
"theta8"]

Y = df["y"]
X = df.drop(columns=["y"])

kf = KFold(n_splits=5, random_state=132, shuffle=True)

clf = MLPRegressor(alpha=0, activation="relu", hidden_layer_sizes=(3, 2),
random_state=1)
clf2 = MLPRegressor(alpha=6, activation="relu", hidden_layer_sizes=(3, 2),
```

```
random_state=1)

res = []
res2 = []
for train_index, test_index in kf.split(X, Y):
    x_train = X.iloc[train_index].loc[:, features].values
    x_test = X.iloc[test_index][features].values
    y_train = Y[train_index]
    y_test = Y[test_index]

    clf.fit(x_train, y_train)
    tmp = list(clf.predict(x_test))

    clf2.fit(x_train, y_train)
    tmp2 = list(clf2.predict(x_test))

    for i in range(0, len(tmp), 1):
        resid = tmp[i] - y_test.iloc[i]
        resid2 = tmp2[i] - y_test.iloc[i]
        res += [resid]
        res2 += [resid2]

print(res)
sns.boxplot(x=res)
plt.title("Without regularization")
plt.show()

print(res2)
sns.boxplot(x=res2)
plt.title("With regularization")
plt.show()
```

END