Aprendizagem 2021/22

Homework III – Group 039

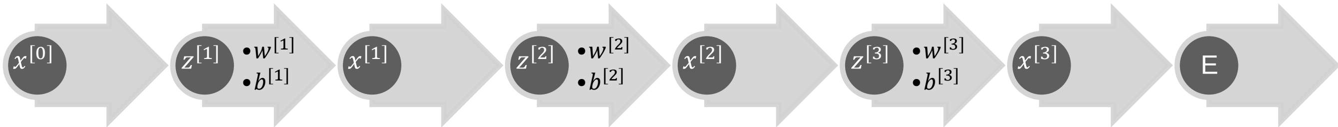
I. Pen-and-paper

1)

a.

$$w^{[1]} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad b^{[1]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad w^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad b^{[2]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad w^{[3]} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad b^{[3]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad x^{[0]} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$t = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad f(x) = \tanh(x) \quad E = MSE \quad \eta = 0.1$$



$$z^{[1]} = w^{[1\}}x^{[0]} + b^{[1]} = \begin{bmatrix} 6 \\ 1 \\ 6 \end{bmatrix} \quad x^{[1]} = f\left(z^{[1]}\right) = \begin{bmatrix} 0.99998771 \\ 0.76159416 \\ 0.99998771 \end{bmatrix}$$

$$z^{[2]} = w^{[2]}x^{[1]} + b^{[2]} = \begin{bmatrix} 3.76156958 \\ 3.76156958 \end{bmatrix} \quad x^{[2]} = f\left(z^{[2]}\right) = \begin{bmatrix} 0.99891972 \\ 0.99891972 \end{bmatrix}$$

$$z^{[3]} = w^{[3]}x^{[2]} + b^{[3]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad x^{[3]} = f\left(z^{[3]}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$w^{[i]\prime} = w^{[i]} - \eta \frac{dE}{dw^{[i]}} \quad b^{[i]\prime} = b^{[i]} - \eta \frac{dt}{db^{[i]}}$$

$$\frac{dx^{[i]}}{dz^{[i]}} = \frac{df(z')}{dz'} = 1 - f\left(z^{[i]}\right)^2 \quad \frac{dz^{[i+1]}}{dx^{[i]}} = \frac{d\left(w^{[i+1]}x^{[i]} + b^{[i+1]}\right)}{dx^{[i]}} = w^{[i+1]}$$

$$\delta^{[3]} = \frac{dE}{dz^{[3]}} = \frac{dE}{dx^{[3]}}\frac{dx^{[3]}}{dz^{[3]}} = \left(x^{[3]} - t\right)\left(1 - f\left(z^{[3]}\right)^2\right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \frac{dE}{dx^{[3]}} = \frac{d}{dx^{[3]}}\frac{1}{2}\left(x^{[3]} - t\right)^2 = x^{[3]} - t$$

$$\delta^{[2]} = \frac{dz^{[3]}}{dx^{[2]}}\delta^{[3]}\frac{dx^{[2]}}{dz^{[2]}} = w^{[3]}\delta^{[3]}\left(1 - f\left(z^{[2]}\right)^2\right) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix}\begin{bmatrix} 0.0021594 \\ 0.0021594 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\delta^{[1]} = \frac{dz^{[2]}}{dx^{[1]}}\delta^{[2]}\frac{dx^{[1]}}{dz^{[1]}} = w^{[2]}\delta^{[2]}\left(1 - f\left(z^{[1]}\right)^2\right) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} 2.45765474 \times 10^{-5} \\ 4.19974342 \times 10^{-1} \\ 2.45765474 \times 10^{-5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\frac{dE}{dw^{[i]}} = \frac{dE}{dz^{[i]}}\frac{dz^{[i]}}{dw^{[i]}} = \delta^{[1]}x^{[i-1]^T} \quad \frac{dE}{db^{[i]}} = \frac{dE}{dz^{[i]}}\frac{dz^{[i]}}{db^{[i]}} = \delta^{[1]}$$

$$w^{[1]\prime} = w^{[1]} - \eta\frac{dE}{dw^{[1]}} = w^{[1]} - \eta\left(\delta^{[1]}x^{[0]^T}\right) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$b^{[1]\prime} = b^{[1]} - \frac{dE}{db^{[1]}} = b^{[1]} - \eta\delta^{[1]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$w^{[2]\prime} = w^{[2]} - \eta\frac{dE}{dw^{[2]}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad b^{[2]'} = b^{[2]} - \eta\frac{dE}{db^{[2]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 039

$$w^{[3]'} = w^{[3]} - \eta \frac{dE}{dw^{[3]}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} -0.09989197 & -0.09989197 \\ -0.09989197 & 0.09989197 \end{bmatrix} = \begin{bmatrix} -0.09989197 & -0.09989197 \\ -0.09989197 & 0.09989197 \end{bmatrix}$$

$$b^{[3]'} = b^{[3]} - \frac{dE}{db^{[3]}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix}$$

b.

$$t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad f(x) = \begin{cases} softmax(x), i = 3 \\ \tanh(x) \end{cases} \quad E = cross\ entropy$$

Same procedure as 1 a until $x^{[3]}$ calculation

$$x^{[3]} = softmax\left(z^{[3]}\right) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$\delta^{[3]} = \frac{dE}{dz^{[3]}} = \left(x^{[3]} - t\right) = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$$

$$\delta^{[2]} = \frac{dz^{[3]}}{dx^{[2]}} \delta^{[3]} \frac{dx^{[2]}}{dz^{[2]}} = w^{[3]} \delta^{[3]} \left(1 - f\left(z^{[2]}\right)^2\right) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \begin{bmatrix} 0.0021594 \\ 0.0021594 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\delta^{[1]} = \frac{dz^{[2]}}{dx^{[1]}} \delta^{[2]} \frac{dx^{[1]}}{dz^{[1]}} = w^{[2]} \delta^{[2]} \left(1 - f\left(z^{[1]}\right)^2\right) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 2.45765474 \times 10^{-5} \\ 4.19974342 \times 10^{-1} \\ 2.45765474 \times 10^{-5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$w^{[1]'} = w^{[1]} - \eta \frac{dE}{dw^{[1]}} = w^{[1]} - \eta \left(\delta^{[1]} x^{[0]^T}\right) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$b^{[1]'} = b^{[1]} - \frac{dE}{db^{[1]}} = b^{[1]} - \eta \delta^{[1]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$w^{[2]'} = w^{[2]} - \eta \frac{dE}{dw^{[2]}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad b^{[2]'} = b^{[2]} - \eta \frac{dE}{db^{[2]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$w^{[3]'} = w^{[3]} - \eta \frac{dE}{dw^{[3]}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} -0.04994599 & -0.04994599 \\ 0.04994599 & 0.04994599 \end{bmatrix} = \begin{bmatrix} 0.04994599 & 0.04994599 \\ -0.04994599 & -0.04994599 \end{bmatrix}$$

$$b^{[3]'} = b^{[3]} - \frac{dE}{db^{[3]}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -0.05 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.05 \\ -0.05 \end{bmatrix}$$

II. Programming and critical analysis

2)

Actual\Predicted	Negative	Positive	
Negative	419	25	*444*
Positive	18	221	*239*
	437	**246**	***683***

Table 1: Confusion Matrix Without Early Stopping

Actual\Predicted	Negative	Positive

Negative	389	55	*444*
Positive	5	234	*239*
	394	**289**	***683***

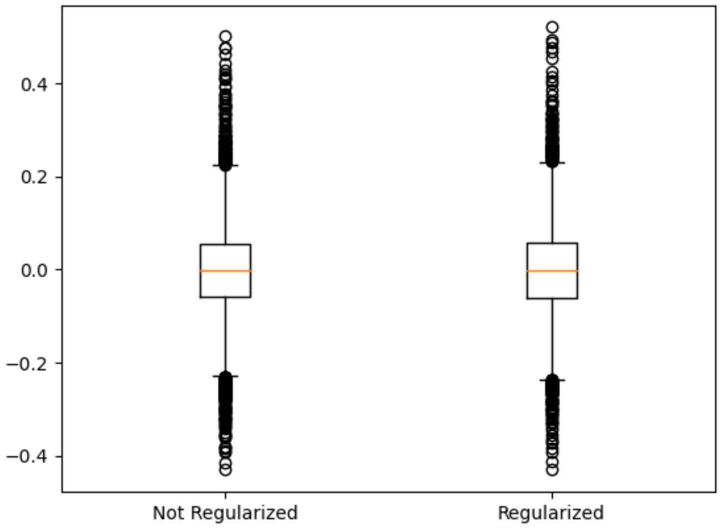
Table 2: Confusion Matrix With Early Stopping

Two reasons for the only slight observed differences are:

1. The use of k-fold cross validation where the model is repeatedly refitted on parts of the dataset.
2. Early stopping being meant to stop a single model when it starts having increased generalized error.

These make them not very suited to be used together

3)



Four strategies that can be used to minimize the observed error of the multi-layer perceptron regressor are:

1. Increase the training sample
2. Early stopping: prevent overfitting by stopping the training when the testing error rate starts increasing.
3. Change the complexity of the network structure and parameters by adding/removing nodes.
4. Regularization: Ensuring the weights keep small, since this indicates a less complex model and therefore more stable and less prone to error from outliers in the input.

III. APPENDIX

```
import pandas as pd

import matplotlib.pyplot as plt
import numpy as np

from scipy.io import arff

from sklearn.neural_network import MLPClassifier
from sklearn.neural_network import MLPRegressor

from sklearn.model_selection import KFold
from sklearn.model_selection import StratifiedKFold

from sklearn.model_selection import cross_val_predict

from sklearn.metrics import confusion_matrix

GROUPN = 0

def quest2():
    # Extract Data
    D_breast = pd.DataFrame( arff.loadarff( "breast.w.arff" )[0] )
    # Elements array
    X = D_breast.drop(columns=D_breast.columns[-1]).to_numpy().astype(int)
    # Results array binarized
    Y = D_breast[D_breast.columns[-1]].replace(b'benign', 0).replace(b'malignant', 1)

    stratifiedk_splits = StratifiedKFold(n_splits=5, random_state=GROUPN,
shuffle=True)

    clf = MLPClassifier(hidden_layer_sizes=(3, 2), random_state=GROUPN)
    Y_pred = cross_val_predict(clf, X, Y, cv=stratifiedk_splits)
    conf_matrix = confusion_matrix(Y, Y_pred)

    clf_es = MLPClassifier(hidden_layer_sizes=(3, 2), random_state=GROUPN,
early_stopping=True)
    Y_es_pred = cross_val_predict(clf_es, X, Y, cv=stratifiedk_splits)
    conf_matrix_es = confusion_matrix(Y, Y_es_pred)

    print("Confusion matrix")
    print(conf_matrix)
    print("Confusion matrix - Early Stopping")
    print(conf_matrix_es)
```

```
def quest3():
    # Extract Data
    D_kin = pd.DataFrame( arff.loadarff( "kin8nm.arff" )[0] )
    # Elements array
    X = D_kin.drop(columns=D_kin.columns[-1]).to_numpy()

    Y = D_kin[D_kin.columns[-1]].to_numpy()

    k_splits = KFold(n_splits=5, random_state=GROUPN, shuffle=True)

    clf = MLPRegressor(alpha=0.1, random_state=GROUPN)
    Y_pred = cross_val_predict(clf, X, Y, cv=k_splits)
    residuals = np.subtract(Y, Y_pred)

    clf_reg = MLPRegressor(alpha=0, random_state=GROUPN)
    Y_reg_pred = cross_val_predict(clf_reg, X, Y, cv=k_splits)
    residuals_reg = np.subtract(Y, Y_reg_pred)

    plt.boxplot([residuals, residuals_reg], labels=("Not Regularized", 
"Regularized"))
    plt.savefig("graph_ex3")

quest2()
quest3()
```

END