Aprendizagem 2021/22

Homework III – Group 114

1. Pen-and-paper

1.

Hw3

1.a) $W^{[1]} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\quad b^{[1]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Activate Function:

$$\tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$$

$W^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ $\quad b^{[2]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$\frac{\partial}{\partial u} \tanh(u) = \dots = 1 - \tanh(u)^2$$

$W^{[3]} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ $\quad b^{[3]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Training: $x = [1\ 1\ 1\ 1\ 1]^T$ $\quad$ target: $t = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Foward Propagation:

$$\vec{z}^{[1]} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \\ 5 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ 6 \end{bmatrix} \qquad \vec{x}^{[1]} = \tanh\left(\begin{bmatrix} 6 \\ 1 \\ 6 \end{bmatrix}\right) = \begin{bmatrix} 0{,}999988 \\ 0{,}761594 \\ 0{,}999988 \end{bmatrix}$$

$$\vec{z}^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0{,}999988 \\ 0{,}761594 \\ 0{,}999988 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2{,}76157 \\ 2{,}76157 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3{,}76157 \\ 3{,}76157 \end{bmatrix} \qquad \vec{x}^{[2]} = \tanh\left(\begin{bmatrix} 3{,}76157 \\ 3{,}76157 \end{bmatrix}\right)$$

$$x^{[2]} = \begin{bmatrix} 0{,}99892 \\ 0{,}99892 \end{bmatrix}$$

$$\vec{z}^{[3]} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0{,}999892 \\ 0{,}999892 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \vec{x}^{[3]} = \tanh\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$E(t, x^{[3]}) = \frac{1}{2} \sum \left(x^{[3]} - t\right)^2 = \frac{1}{2}(1 + 1) = 1$$

$$\frac{\partial E}{\partial x^{[l]}}(t, x^{[l]}) = x^{[l]} - t \qquad \frac{\partial z^{[l]}}{\partial b^{[l]}}(w^{[l]}, b^{[l]}, x^{[l-1]}) = 1$$

$$\frac{\partial x^{[l]}}{\partial z^{[l]}}(z^{[l]}) = (1 - \tanh(x)^2) \qquad \frac{\partial z^{[l]}}{\partial x^{[l-1]}}(w^{[l]}, b^{[l]}, x^{[l-1]}) = w^{[l]}$$

$$\frac{\partial z^{[l]}}{\partial w^{[l]}} = (w^{[l]}, b^{[l]}, x^{[l-1]}) = x^{[l-1]}$$

$$\delta^{[3]} = \frac{\partial E}{\partial x^{[3]}} \circ \frac{\partial x^{[3]}}{\partial z^{[3]}} = (x^{[3]} - t) \circ \left(1 - \tanh(z^{[3]})^2\right) = \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) \circ \left(\left(1 - \begin{bmatrix} 0 \\ 0 \end{bmatrix}^2\right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right.$$

$$\delta^{[2]} = \frac{\partial z^{[3]}}{\partial x^{[2]}}^T \cdot \delta^{[3]} \cdot \frac{\partial x^{[2]}}{\partial z^{[2]}} = \left[W^{[3]}\right]^T \cdot \delta^{[3]} \cdot \left(1 - \tanh(z^{[2]})^2\right) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} \cdot \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0{,}9999892 \\ 0{,}9999892 \end{bmatrix}^2\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\delta^{[1]} = \frac{\partial z^{[2]}}{\partial x^{[1]}}^T \cdot \delta^{[2]} \cdot \frac{\partial x^{[1]}}{\partial z^{[1]}} = (w^{[2]})^T \cdot \delta^{[2]} \cdot \left(1 - \tanh(z^{[1]})^2\right) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \left[\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{bmatrix} 0{,}999988 \\ 0{,}761594 \\ 0{,}999988 \end{bmatrix}^2\right] = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}} = \delta^{[1]} \cdot \frac{\partial z^{[1]}}{\partial w^{[1]}}^T = \delta^{[1]} \cdot (x^{[0]})^T = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \cdot [1\ 1\ 1\ 1\ 1] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 114

$$w^{[1]} = w^{[1]} - \eta \frac{\partial E}{\partial w^{[1]}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta^{[1]} \cdot \frac{\partial z^{[1]}}{\partial b^{[1]}}^T = \delta^{[1]} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad b^{[1]} = b^{[1]} - \eta \frac{\partial E}{\partial b^{[1]}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[2]}} = \delta^{[2]} \cdot \frac{\partial z^{[2]}}{\partial b^{[2]}}^T = \delta^{[2]} \cdot \left[x^{[1]}\right]^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0{,}999988 & 0{,}761594 & 0{,}999988 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$w^{[2]} = w^{[2]} - \eta \frac{\partial E}{\partial w^{[2]}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \delta^{[2]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad b^{[2]} = b^{[2]} - \eta \frac{\partial E}{\partial b^{[2]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[3]}} = \delta^{[3]} \cdot \frac{\partial z^{[3]}}{\partial b^{[3]}}^T = \delta^{[3]} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0{,}99892 & 0{,}99892 \end{bmatrix} = \begin{bmatrix} -0{,}99892 & -0{,}99892 \\ 0{,}99892 & 0{,}99892 \end{bmatrix}$$

$$b^{[3]} = b^{[3]} - \eta \frac{\partial E}{\partial b^{[3]}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0{,}1 \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 0{,}1 \\ -0{,}1 \end{bmatrix}$$

$$w^{[3]} = w^{[3]} - \eta \frac{\partial E}{\partial w^{[3]}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}9999892 & -0{,}9999892 \\ 0{,}9999892 & 0{,}9999892 \end{bmatrix} = \begin{bmatrix} 0{,}09999892 & 0{,}09999892 \\ -0{,}09999892 & -0{,}09999892 \end{bmatrix}$$

b) training = $x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$ $\qquad z = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\qquad \eta = 0{,}1$

$$\mathrm{softmax}\left([z_1 \ldots z_n]^T\right) = [u_1 \ldots u_n]^T \qquad u_i = \frac{\exp(z_i)}{\sum_k \exp(z_k)} \qquad E(t, x^{[3]}) = -\sum_{i} t_i \log x_i^{[3]}$$

$$i = j \Rightarrow \frac{\partial u_i}{\partial z_j} = \frac{\partial}{\partial z_j} \frac{\exp(z_i)}{\sum_{k=1} \exp(z_k)} = u_i(1 - u_j) \qquad \frac{\partial x^{\ell}}{\partial z^{\ell}}(z^{[\ell]}) = 1 - \tanh(z^{[\ell]})^2$$

$$i \neq j \Rightarrow \frac{\partial u_i}{\partial z_j} = -u_i \cdot u_j \qquad \delta^{[3]} = \begin{bmatrix} \delta_1^3 \\ \delta_2^3 \\ \delta_3^3 \end{bmatrix} = \begin{bmatrix} x_1^3 - t_1 \\ x_2^3 - t_2 \\ x_3^3 - t_3 \end{bmatrix} \qquad \frac{\partial z^{\ell}}{\partial w^{\ell}}(w^{[\ell]}, b^{[\ell]}, x^{[\ell-1]}) = x^{[\ell-1]}$$

$$z^{[3]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad , \quad x^{[3]} = \mathrm{softmax} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\exp(0)}{2\exp(0)} \\ \frac{\exp(0)}{2\exp(0)} \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \qquad \frac{\partial z^{[\ell]}}{\partial b^{[\ell]}}(w^{[\ell]}, b^{[\ell]}, x^{[\ell]}) = 1$$

$$\delta^{[3]} = \begin{bmatrix} 0{,}50 \\ 0{,}50 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix} \qquad \frac{\partial z^{[\ell]}}{\partial x^{[\ell-1]}}(w^{[\ell]}, b^{[\ell]}, x^{[\ell-1]}) = w^{[\ell]}$$

$$\delta^{[2]} = \begin{bmatrix} 0.\text{[redacted]} & 0.\text{[redacted]} \\ 0.\text{[redacted]} & 0.\text{[redacted]} \end{bmatrix} \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix} \cdot \begin{pmatrix} 1 - (0{,}99982)^2 \\ 1 - (0{,}99982)^2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\delta^{[1]} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \left(1 - \begin{bmatrix} 0{,}999988 \\ 0{,}761594 \\ 0{,}999988 \end{bmatrix}^2 \right) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$w^{[1]} = w^{[1]} - \eta \frac{\partial E}{\partial w^{[1]}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[1]}} = \delta^{[1]} \cdot \frac{\partial z^{[1]}}{\partial w^{[1]}}^T = \delta^{[1]} \cdot (x^{[0]})^T = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \frac{\partial E}{\partial b^{[1]}} = \delta^1 \cdot \frac{\partial z^{1T}}{\partial b^1} = \delta^{[1]} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$b^{[1]} = b^{[1]} - \eta \frac{\partial E}{\partial b^{[1]}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial w^{[2]}} = \delta^{[2]} \cdot \frac{\partial z^{[2]}}{\partial w^{[2]}}^T = \delta^{[2]} \cdot (x^{[1]})^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot (0{,}999988 \;\; 0{,}761594 \;\; 0{,}999988) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$w^{[2]} = w^{[2]} - \eta \frac{\partial E}{\partial w^{[2]}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \frac{\partial E}{\partial b^{2}} = \delta^2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad b^{[2]} = b^{[2]} - \eta \frac{\partial E}{\partial b^{[2]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Aprendizagem 2021/22

Homework III – Group 114

$$\frac{\partial E}{\partial w^3} = \delta^{[3]} \cdot \frac{\partial z^{[3]}}{\partial w^{[3]}}^T = \delta^{[3]} \cdot \left(x^{[2]}\right)^T = \overset{2\times1}{\begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix}} \cdot \overset{1\times2}{\begin{pmatrix} 0{,}999892 & 0{,}999892 \end{pmatrix}} = \begin{bmatrix} -0{,}499946 & -0{,}499946 \\ 0{,}499946 & 0{,}499946 \end{bmatrix}$$

$$w^3 = w^3 - \eta \frac{\partial E}{\partial w^3} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}499946 & -0{,}499946 \\ 0{,}499946 & 0{,}499946 \end{bmatrix} = \begin{bmatrix} 0{,}0499946 & 0{,}0499946 \\ -0{,}0499946 & -0{,}0499946 \end{bmatrix}$$

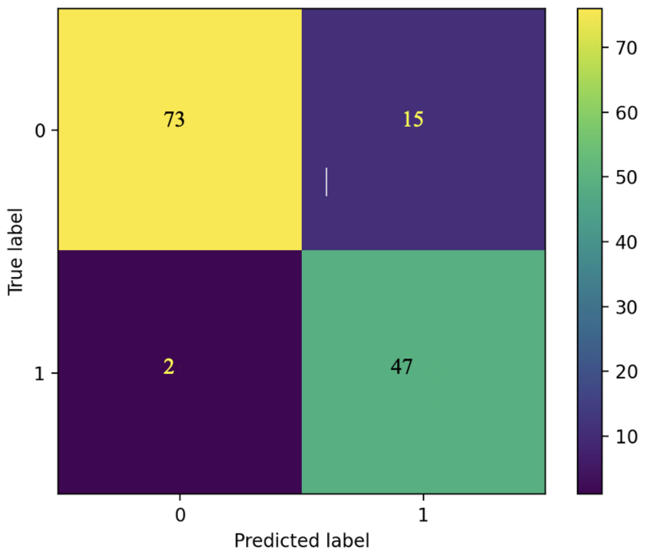
$$\frac{\partial E}{\partial b^3} = \delta^{[3]} \cdot \frac{\partial z^{[3]}}{\partial b^{[3]}}^T = \delta^{[3]} = \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix}$$

$$b^3 = b^3 - \eta \frac{\partial E}{\partial b^3} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix} = \begin{bmatrix} 0{,}05 \\ -0{,}05 \end{bmatrix}.$$

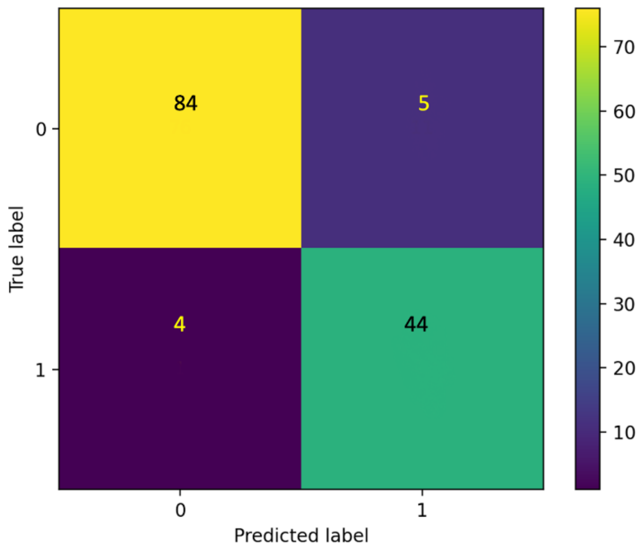
Os passos intermédios, nomeadamente cálculo de derivadas, foram baseados no documento pdf disponibilizado pelo Prof. Andrzej Wichert em pratical lectures.

II. Programming and critical analysis

2. Answer 2



(early stopping = false)



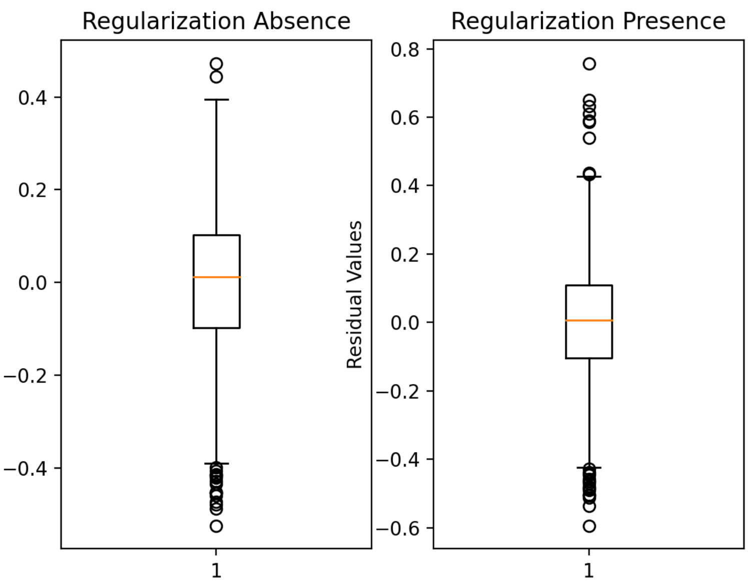
(early stopping = true)

É possível reduzir o overfitting através da redução do espaço de amostra de dados, dividindo a amostra em menos quantidades, e também fixando o tamanho de cada dimensão do espaço permitindo que o erro fique reduzido ao máximo. Assim que o erro começa a aumentar, o early stopping termina a classificação, o que é positivamente verificável nesta experiência.

Aprendizagem 2021/22

Homework III – Group 114

3. Answer 3



Para reduzir o erro é necessário:

- ☐ fixar a regularização,entre 0 e 1, aumentando o fator de aprendizagem;
- ☐ fixar o random_state diferente de 0 (por exemplo 114) (shuffle);
- ☐ Aumentar o número de iterações(ephocs), quanto mais iterações mais pequeno vai ser o erro.
- ☐ Alterar a função de ativação : tanh(x)

Aprendizagem 2021/22

Homework III – Group 114

III. APPENDIX

Exercise 2

```
import pandas as pd
import numpy as np
from scipy.io.arff.arffread import print_attribute
from scipy.io.arff import loadarff
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.neural_network import MLPClassifier
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import plot_confusion_matrix
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
from sklearn.neural_network import MLPRegressor
from sklearn.model_selection import cross_val_predict

if __name__ == '__main__':
    raw_data = loadarff('breast.w.arff')
    df_data = pd.DataFrame(raw_data[0])
    classe = df_data.pop('Class')
    df_data = df_data.values
    df_data = df_data.astype(int)

    Y = np.array(classe)
    Y = Y.astype(int)
kf = KFold(n_splits=5,random_state=0,shuffle=True)

clf = MLPClassifier(hidden_layer_sizes=(3,2),early_stopping=False,random_state=120)
#clf = MLPClassifier(hidden_layer_sizes=(3,2),early_stopping=True,random_state=120)
for train_ind, test_ind in kf.split(df_data):
    X_train, X_test = df_data[train_ind], df_data[test_ind]
    y_train, y_test = Y[train_ind], Y[test_ind]
    clf.fit(X_train,y_train)
    plot_confusion_matrix(clf,X_test,y_test)
    plt.show()
```

Exercise 3

```
if __name__ == '__main__':
    raw_data = loadarff('kin8nm.arff')
    df_data = pd.DataFrame(raw_data[0])
    Y = df_data.pop('y')
    df_data = df_data.values
    df_data = df_data.astype(float)
alphas = np.logspace(-1, 1, 5)
kf = KFold(n_splits=5,random_state=0,shuffle=True)
vec = []
fig, axs = plt.subplots(1,2)
for train_ind, test_ind in kf.split(df_data):
    X_train, X_test = df_data[train_ind], df_data[test_ind]
    y_train, y_test = Y[train_ind], Y[test_ind]
    regr = MLPRegressor(hidden_layer_sizes=(3,2),activation='relu',random_state=
0,alpha=0).fit(X_train, y_train)
    y_test_data_pred = regr.predict(X_test)
    fold_testing_error = y_test-y_test_data_pred
vec.append(fold_testing_error)
axs[0].boxplot(vec)
axs[0].set_title("Regularization Absence")
kf = KFold(n_splits=5,random_state=0,shuffle=True)
```

```
vet=[]
for train_ind, test_ind in kf.split(df_data):
    X_train, X_test = df_data[train_ind], df_data[test_ind]
    y_train, y_test = Y[train_ind], Y[test_ind]
    regr = MLPRegressor(hidden_layer_sizes=(3,2),activation='relu',random_state=
0,alpha=1).fit(X_train, y_train)
    y_test_data_pred = regr.predict(X_test)
    fold_testing_errors = y_test-y_test_data_pred
vet.append(fold_testing_errors)
axs[1].boxplot(vet)
axs[1].set_title("Regularization Presence")
plt.ylabel("Residual Values")
plt.show()
```

END