**TP03**

**Problématique:**

On voudrait faire une classification Binaire (on possède deux classes): Dans la première expérience, on dispose d’une base de données (table) qui contient des fruits avec différents features le nombre de fruits est 4 : apple (pomme), mandarin, Orange, Lemon.

Puisque on procède à une classification binaire , nous allons avoir donc deux classes :

1. Classe Apple
2. Classe autre

l’objectif est de classer les fruits selon deux features height and width dans la classe Apple ou la classe autre.

Dans la deuxième expérience , nous allons utiliser une base réelle : Cancer .

L’objectif est de classer les patients selon leur features en classe « Benin » ou « Malin ».

Ce TP est en partie donné , veuillez le complétez en écrivant le code qui convient à la place des pointillés .

**Etape 1:**

Importez les Librairies nécessaires (plot, pandas, split, ..) en plus de ces deux :

……..

………

……..

Algorithme de la Régression logistique :

from sklearn.linear\_model import LogisticRegression

Utilitaires de plot pour la classification :

from adspy\_shared\_utilities import (plot\_class\_regions\_for\_classifier\_subplot)

**Etape 2:**

Importez la base de données qui s’appelle : fruit\_data\_with\_colors.txt :

fruits = pd.read\_table(……………..')

Remarque :

Dans la base , les noms des features (caractéristiques/variables) X sont:

feature\_names\_fruits = ['height', 'width', 'mass', 'color\_score']

Les labels des fruits de la base sont :

target\_names\_fruits = ['apple', 'mandarin', 'orange', 'lemon']

On récupère deux caractéristiques de fruits : height et width:

X\_fruits\_2d = fruits[['height', 'width']]

y\_fruits\_2d = fruits['fruit\_label']

TOUTE LA PARTIE QUI SUIT EST OPTIONELLE, ELLE SERT A DESSINER UN GRAPHIQUE QUI MONTRE LA CLASSIFICATION VOUS POUVEZ IGNORER JUSQU A CE QUE VOUS TERMINEZ VOTRE TP.

fig, subaxes = plt.subplots(1, 1, figsize=(7, 5))

plot\_class\_regions\_for\_classifier\_subplot(clf, X\_train, y\_train, None,

None, 'Logistic regression \

for binary classification\nFruit dataset: Apple vs others',

subaxes)

subaxes.set\_xlabel('height')

subaxes.set\_ylabel('width')

ICI SE TERMINE LA PARTIE OPTIONELLE.

On veut faire une classification de pomme Apple contre tous :

* Classification binaire(deux classes)

y\_fruits\_apple = y\_fruits\_2d == 1 # make into a binary problem: apples vs everything else

**Etape 3:**

Divisez la base en base train et base test : si warning est affiché alors remplacez .as\_matrix par .values

….., ……, ……, …….. = …………(………….., y\_fruits\_apple.values, random\_state = …..)

**Etape 4:**

Appelez le classifieur régression logistique :

clf = ……….().fit(……., ……..)

**Etape 5:**

prédire la classe d’un fruit qui a un h(height)=6 et w(weight)=8 , en utilisant clf.predict()

et affichez le :

h = 6

w = 8

print('A fruit with height {} and width {} is predicted to be: {}'

.format(h,w, ['not an apple', 'an apple'][clf.predict([[h,w]])[0]]))

prédire la classe d’un fruit qui a un h(height)=10 et w(weight)=7 , en utilisant clf.predict(….)

et affichez le :

h = 10

w = 7

……………………………………………………………………………………………………………………………………….

**Etape 6:**

imprimez le score train et test de votre classification :

………('Accuracy of Logistic regression classifier on training set: {:.2f}'

.format(clf……….(………, ………)))

print('Accuracy of Logistic regression classifier on test set: {:.2f}'

.format(clf………(………., ………..)))

Question : Qu’est ce que vous remarquez?

# #### Application to real dataset

**Etape 1 :**

Depuis sklearn.datasets Importez la base réelle load\_breast\_cancer:

from sklearn.datasets import cancer = load\_breast\_cancer()

(X\_cancer, y\_cancer) = load\_breast\_cancer(return\_X\_y = True)

**Etape 2:**

Divisez la base en train et test:

………………………………………………………………………………………………………………………………

**Etape 3:**

Appelez l’algorithme de la régression logistique en donnant vos paramètres

………………………………………………………………………………………………………………….

**Etape 4:**

imprimez les scores train et test :

print('Breast cancer dataset')

…………………………………………………………………..

…………………………………………………………………..

Question: Qu’est ce que vous remarquez?

ICI SE TERMINE VOTRE TP. FIN.

CE QUI SUIT EST EN PLUS VOUS N ETES PAS OBLIGE DE LE VOIR POUR LE MOMENT ;

# #### Logistic regression regularization: C parameter

# In[18]:

X\_train, X\_test, y\_train, y\_test = (train\_test\_split(X\_fruits\_2d.as\_matrix(),

y\_fruits\_apple.as\_matrix(), random\_state=0))

fig, subaxes = plt.subplots(3, 1, figsize=(4, 10))

for this\_C, subplot in zip([0.1, 1, 100], subaxes):

clf = LogisticRegression(C=this\_C).fit(X\_train, y\_train)

title ='Logistic regression (apple vs rest), C = {:.3f}'.format(this\_C)

plot\_class\_regions\_for\_classifier\_subplot(clf, X\_train, y\_train,

X\_test, y\_test, title, subplot)

plt.tight\_layout()