

## Projet 1 –TrD

### Les objectifs du projet :

Mettre en place des classifieurs de Bayes et le classifient des lus proches voisins (kppv) sur un jeu de données réelles

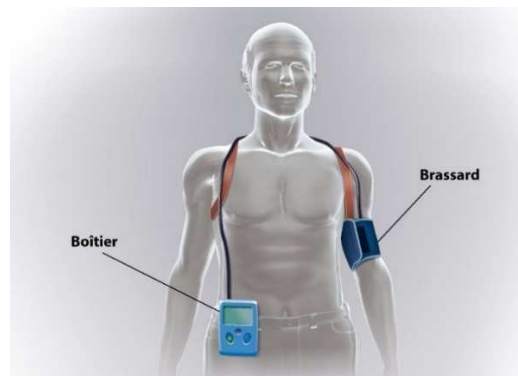
Ce projet est repris des projets proposés par S. Charbonnier dans le cadre du cours de Machine Learning des IESE5.

---

## Holter tensionnel intégrant la détection d'artéfacts de mouvement

### 1. Contexte et objectifs

Afin d'établir un diagnostic d'hypertension artérielle, la simple mesure de la tension artérielle au cabinet d'un médecin ne suffit souvent pas. En effet, la tension artérielle varie au cours de la journée en fonction de différents facteurs dont le stress, l'activité physique, l'heure de la journée ... Afin d'établir un diagnostic fiable, une solution est d'équiper le patient d'un Holter tensionnel (Figure 1). Cet appareil mesure de manière automatique la tension artérielle du patient pendant 24 heures, en réalisant une mesure toutes les 15 minutes en journée et toutes les 30 minutes la nuit. La mesure automatique de la pression artérielle est réalisée par méthode oscillométrique.

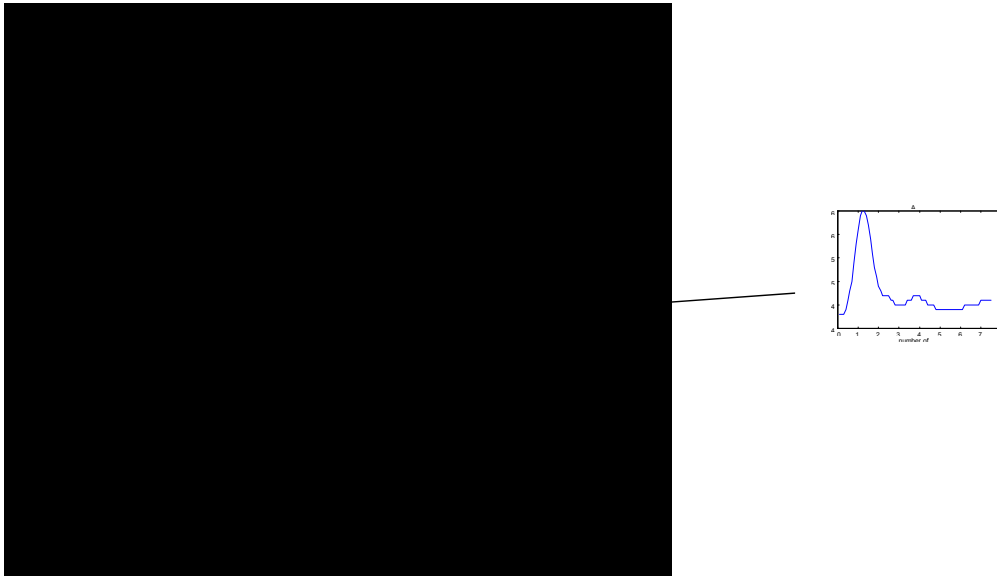


*Figure 1 :* Patient équipé d'un Holter tensionnel

### **Principe de la mesure de pression artérielle par méthode oscillométrique**

Cette méthode permet la prise automatique de la mesure de la pression artérielle, sans l'intervention d'un soignant. Le patient est équipé d'un brassard, qui, lors de la mesure, va se gonfler à une pression supérieure à la pression systolique puis se dégonfler pas à pas jusqu'à une pression inférieure à la pression diastolique. A chaque pas de pression, la variation de pression dans le brassard générée par la circulation du sang est mesurée à l'aide d'un capteur de pression. Celle-ci a une forme pulsée. La forme générée est appelée onde de pouls (Figure 2).

A chaque pas de pression, l'onde de pouls est enregistrée et son amplitude est mesurée. A la fin de la mesure, une fois le brassard dégonflé, la courbe des amplitudes est analysée et les pressions diastolique, systolique et moyenne sont calculées à partir de cette courbe.



**Figure 2 :** Principe de la mesure de pression par méthode oscillométrique. Figure du haut : pression appliquée par le système au brassard, figure du milieu : mesure de pression dans le brassard avec un zoom sur une onde de pouls, figure du bas : son mesuré au stéthoscope (à titre comparatif avec la mesure oscillométrique).

### **Objectif du système à développer**

Un inconvénient de cette méthode est sa sensibilité aux mouvements du patient. Si le patient bouge le bras au cours de la mesure, l'onde de pouls est déformée, l'amplitude mesurée est erronée, ce qui peut conduire au final à une erreur de mesure sur les pressions diastoliques et systoliques. Une solution éventuelle pour rendre le système plus robuste aux artefacts de mouvements serait d'inclure un système de détection capable de reconnaître en ligne un artefact de mouvement. Si un artefact est détecté, le système peut maintenir le palier de pression plus longtemps de manière à faire l'acquisition d'une nouvelle onde de pouls.

**L'objectif du projet est d'évaluer l'intérêt d'intégrer au holter tensionnel un système de décision capable de trier les artefacts des ondes de pouls, en ligne.** A chaque acquisition d'une onde de pouls, le système détermine si l'onde de pouls est correcte ou si c'est un artefact. Dans le cas d'un artefact, le système de mesure peut refaire une acquisition pour le même palier de pression.

La détection d'artefact en ligne se fait à l'aide d'un classifieur appris à partir de données d'apprentissage. Les caractéristiques qui pourraient permettre de reconnaître un artefact ont déjà été choisies (elles sont décrites plus bas).

### **L'objectif du projet est de :**

- 1/ Construire le système de décision automatique
- 2/ Evaluer ses performances
- 3/ Discuter des résultats et de l'intérêt de son intégration dans le système existant.

Le bilan de ce travail doit présenter et discuter les choix qui ont été faits pour le système de décision, en prenant en compte des critères de performances mais aussi d'utilisabilité pour le patient, pour le cardiologue, etc. Une analyse critique des caractéristiques choisies sera aussi la bienvenue. Les choix doivent être argumentés non pas avec des intuitions mais avec des analyses et des mesures chiffrées.

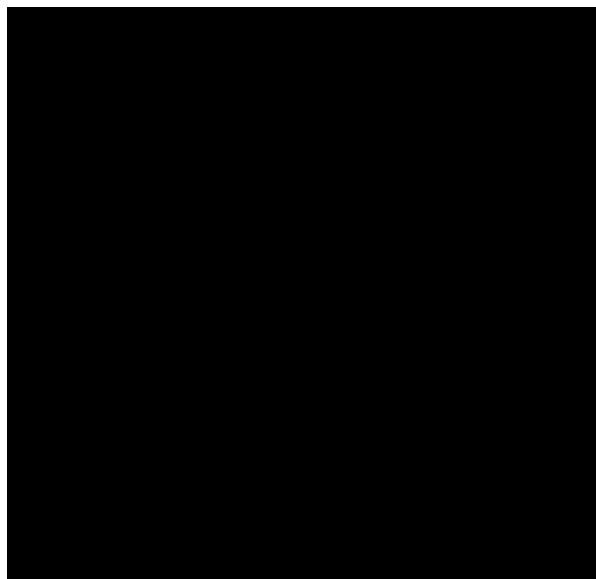
## 2. Elaboration du détecteur

### **Caractéristiques proposées :**

Comme la forme de l'onde de pouls varie d'un patient à un autre et également en fonction de la pression dans le brassard, le système de décision compare l'onde de pouls à analyser à une onde de référence. Lors de l'installation du holter sur le patient par le cardiologue, le système fait l'acquisition d'ondes de pouls de référence, enregistrées à des différentes valeurs de pression ; 8 formes d'ondes sont alors stockées.

Le système de détection opère de la façon suivante. Dès l'acquisition d'une nouvelle onde de pouls à analyser, la valeur maximale de la fonction d'inter-corrélation entre le pulse à analyser et chacun des 8 pulses de référence est calculée. Le pulse de référence pour lequel la corrélation est la plus forte est sélectionné. Le système calcule ensuite la différence entre le pulse de référence et le pulse à analyser de la valeur maximale, la valeur minimale et la largeur. Il calcule aussi la différence entre la pression dans le brassard correspondant au pulse à analyser et au pulse de référence.

Chaque pulse est ainsi décrit par 5 caractéristiques : le coefficient d'inter-corrélation, la différence entre les valeurs maximum, la différence entre les valeurs minimum, la différence de pression dans le brassard et la différence de largeur, (Figure 3). Ces caractéristiques correspondent à ce que l'on appelle le « vecteur forme » du pulse et c'est à partir de ce vecteur forme que le pulse sera classé comme correct ou artéfact.



*Figure 3 :* Onde de pouls de forme correct. La valeur maximale (max), la valeur minimale (min) et la largeur du pulse (width) sont indiqués. Ces 3 grandeurs extraites sur l'onde à analyser sont

comparées à celles extraites sur une onde de référence (celle qui a l'inter corrélation la plus forte avec l'onde à analyser)

### **Base de données :**

Nous disposons d'une base de données composée de trois enregistrements (fichier data.mat), 2 enregistrements ont été effectués sur un premier sujet (matrices data1 et data2), un troisième enregistrement effectué sur un deuxième sujet (matrice data 3). Chaque forme d'onde composant la base de données a été analysée par un médecin cardiologue et rangée suivant 2 classes

1 : artéfact

2 : forme correcte

Lors des enregistrements, les 2 sujets ont volontairement bougé le bras ou les doigts, de manière à générer des artéfacts. Le nombre de pulses artéfactés présent ne traduit donc pas ce que l'on pourrait observer sur un appareil en utilisation normale.

Les matrices data\* sont composées comme suit :

Colonne 1 – numéro du pulse

Colonne 2 – classement (artéfact ou correct) donné par le cardiologue

Colonne 3 – instant correspondant au début du pulse

Colonne 4 – instant correspondant à la fin du pulse

Colonne 5 – maximum de la fonction d'intercorrélation

Colonne 6 – différence entre valeurs maximales

Colonne 7 – différence entre valeurs minimales

Colonne 8 – différence entre les pressions du brassard

Colonne 9 – différence entre les largeurs

A titre d'information, les enregistrements de la pression dans le brassard et donc, les formes d'ondes sont consultables. Ils se trouvent dans les fichiers pulse\*.ASC. Les signaux sont échantillonnés à 100Hz.

Les fichiers pulses\*.ASC sont composés de 3 colonnes.

Colonne 1 : instant d'échantillonnage

Colonne 2 : variation de la pression (ondes de pouls)

Colonne 3 : pression dans le brassard

Les ondes de pouls de la matrice data1 se trouvent dans les fichiers pulse11, pulse12, pulse13.

Chaque onde de pouls j peut être visualisée par la commande :

```
plt.plot(pulse11[data1[j-1,3]:data1[j-1,4],1])
```

Pour la matrice data1,

les pulses de 1 à 69	se trouvent dans le fichier	pulse11
----------------------	-----------------------------	---------

les pulses de 70 à 134	se trouvent dans le fichier	pulse12
------------------------	-----------------------------	---------

135 à la fin		pulse13
--------------	--	---------

Pour la matrice data 2

Les pulses	1 à 80	“	pulse21
Les pulses	81 à 168	“	pulse22
Les pulses	169 à la fin	“	pulse 23

Pour la matrice data3

Les pulses	1 à 75	“	pulse31
Les pulses	76 à 142	“	pulse32
Les pulses	143 à la fin	“	pulse33

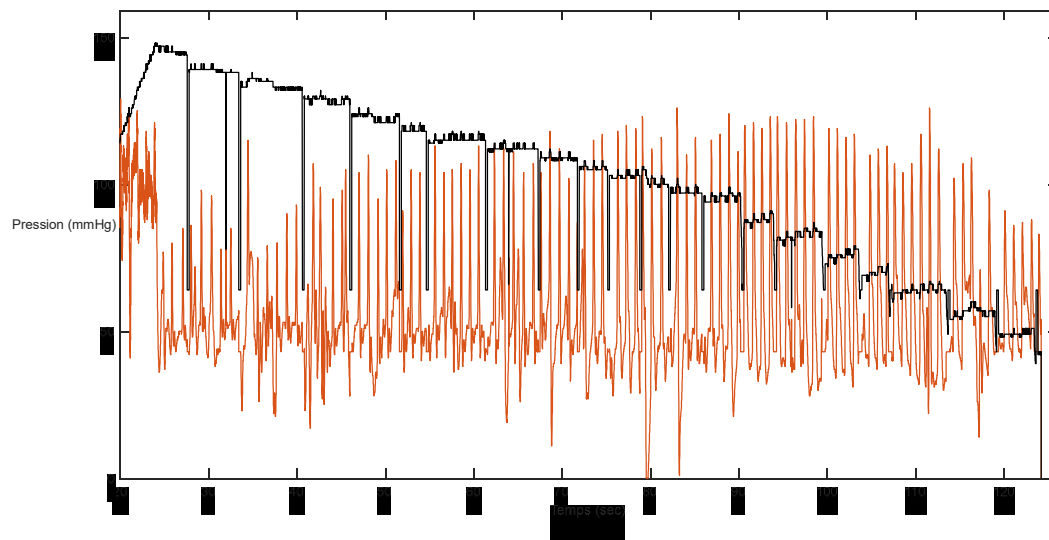


Figure 4 : Pression appliquée (en noir) et pression dans le brassard (en marron) – fichier pulse11

### 3. Quelques pistes de travail, pour répondre au problème

#### 3.1. Récupérer les données

```
# Pour charger des données au format. mat
import scipy.io
data = scipy.io.loadmat('data.mat')
```

Créer un tableau (dataframe ou array) pour jeu de données : data1, data2 et data3.

#### 3.2. Décrire chaque jeu de données

- indiquer le nombre de pulses et le nombre d'artefacts.
- réaliser une ACP pour visualiser les données et afficher les pulses et les artefacts avec des couleurs différentes.
- réaliser également une ALD selon une approche descriptive

#### 3.3. Mettre en place des classifieurs

L'apprentissage sera ici vu comme un problème de classification à 2 classes, les pulses corrects et les artefacts.

Vous testerez 4 classifieurs : le classifieur linéaire, le classifieur quadratique, le classifieur bayésien naïf et les k plus proches voisins.

- Vous utiliserez data1 pour l'apprentissage puis data2 pour le test (intra sujet)
- Vous utiliserez data1 pour l'apprentissage puis data3 pour le test (inter sujet)

### 3.3. Evaluer les performances des classifieurs et les comparer

Pour l'ensemble des classifieurs implémentés, évaluer les performances à l'aide de la matrice de confusion. En déduire le pourcentage de bien classés, le taux de faux positifs et de faux négatifs.

Une « réelle » classification inter sujet se fait en mélangeant les données par exemple ici des 2 patients ; puis une partie de données en utiliser en apprentissage et l'autre partie en test.

**Remarques :** classer une forme correcte en artéfact aura comme conséquence d'augmenter le temps de mesure. Le système de mesure devra attendre de faire l'acquisition d'une nouvelle onde de pouls pour en calculer l'amplitude. Or, la mesure ne doit pas excéder un temps donné car la compression de l'artère par le brassard devient douloureuse pour le patient. Il peut être judicieux d'appliquer des coûts différents en fonction des erreurs commises par le système (fausse alarme, non détection).

Pour chaque stratégie d'apprentissage, il faut penser aux conséquences sur le produit final et notamment au travail supplémentaire pour le cardiologue. Ceci doit rentrer en compte dans le choix du système final retenu.

## 4. Librairies et fonctions utiles

```
#Pour charger des données au format .mat import scipy.io
data = scipy.io.loadmat('data.mat')
```

```
#Pour centrer et réduire des données from sklearn import preprocessing
scaler = preprocessing.StandardScaler()
```

```
# Classe pour l'ACP from sklearn.decomposition import PCA
# Instanciation
acp = PCA()
#affichage des paramètres print(acp)
# On récupère les coordonnées des observations dans le nouvel espace
coord = acp.fit_transform(X)
# variance expliquée
print(acp.explained_variance_ratio_)
```

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
lda = LinearDiscriminantAnalysis()
# X contient les données avec en ligne les individus et en colonne les caractéristiques et y les
# classes des individus
X_r2 = lda.fit(X, y).transform(X)
#
print(lda.explained_variance_ratio_)
```

On utilisera également plt.scatter pour visualiser les points-individus

```
# les classifieurs
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import plot_confusion_matrix, confusion_matrix

from sklearn.model_selection import train_test_split
```