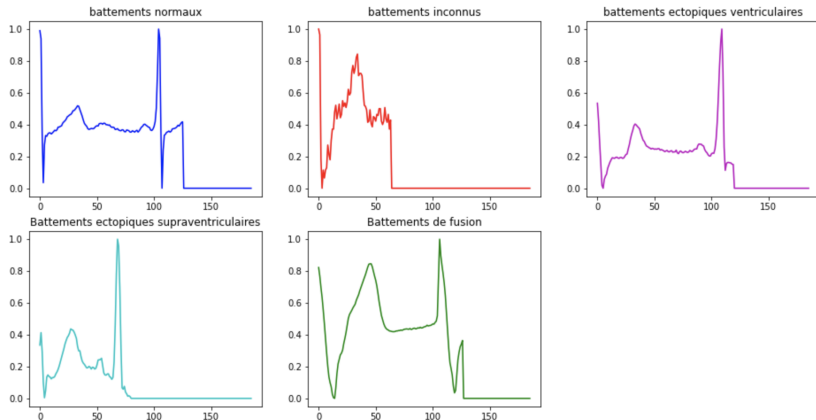


**Projet Analyse de données :**

## Classification du rythme cardiaque à partir de mesures ECG

3A SN : Analyse de données et classification



**Figure:** Les 5 natures de battements à détecter : battements normaux, battements inconnus, battements ectopiques ventriculaires, Battements ectopiques supraventriculaires, Battements de fusion

## Objectif

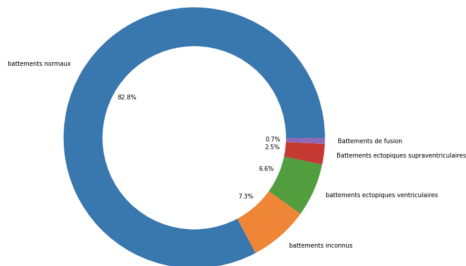
Prédire la nature du battement cardiaque en utilisant l'ECG.

Les signaux correspondent à des formes d'électrocardiogramme (ECG) de battements cardiaques pour le cas normal et les cas affectés par différentes arythmies et infarctus du myocarde.

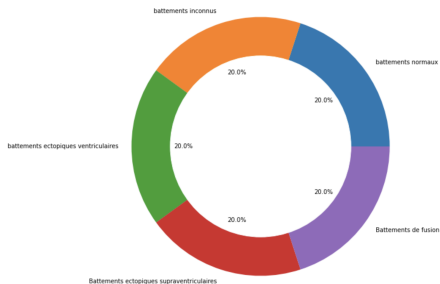
Il est fourni pour chaque enregistrement de l'ensemble de données suivant :

- Le segment du signal ECG.
- Son étiquette d'activité.
- Un identifiant du sujet qui a réalisé l'expérience.

Ces signaux sont prétraités et segmentés, chaque segment correspondant à un battement cardiaque.



Données brutes



Données rééquilibrées

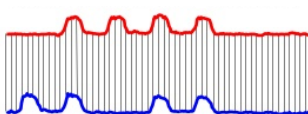
## Organisation :

Ce projet en 6 séances se décomposent en 5 parties :

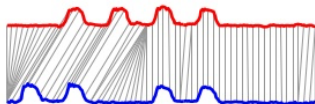
- *Partie I* : Système de reconnaissance d'activité physique avec la DTW
- *Partie II* : Réduction de dimension par ACP et classification par kppv
- *Partie III* : Classification par Forêts aléatoires
- *Partie IV* : Classification par réseaux de neurones
- *Partie V* : Votre étude

## Projet :

- **Langage** : Notebook Python
  - Mise à disposition d'un tutoriel python sous moodle
  - Possibilité de partager le notebook via *Google Colab*
- Travail à réaliser **en binôme**
- **Livrables du projet** : un notebook par binôme et un rapport au format pdf de 10 pages max.
- **Deadline** : le 5 février !



Distance Euclidienne



Dynamic Time Warping (DTW)

⇒ Utilisation de la DTW, algorithme de déformation temporelle dynamique pour synchroniser et aligner des séries temporelles entre elles

⇒ Implémenter l'algorithme de DTW.

### ① Utilisation de la DTW pour la classification supervisée

Soit un dictionnaire  $\{R_1, R_2, \dots, R_N\}$  constitué des exemples de séquences qui sont connues à l'avance (*base d'apprentissage*).

L'algorithme va consister à rechercher la référence  $R_m$  la plus proche d'une séquence dans la *base de test*  $M$  à identifier à l'aide d'une distance  $D$  :

$$m = \arg_{1 \leq i \leq N} \min (\text{Score} (M, R_i))$$

### ② Evaluation de la DTW par matrice de confusion et pourcentage d'éléments bien classés:

fonctions respectives *confusion\_matrix* et *accuracy\_score* de la librairie python *scikit-learn*

**Matrice de confusion :** consiste à compter le nombre de fois où des observations de la classe  $A$  ont été rangées dans la classe  $B$ .

- Chaque ligne de la matrice de confusion représente la *classe réelle* tandis que chaque colonne représente une *classe prédite*.
- Les éléments diagonaux représenteront le nombre d'éléments bien classés. *Classification parfaite = matrice de confusion diagonale*
- Les éléments hors diagonaux représenteront les erreurs de classification.

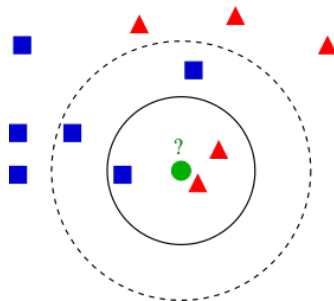


*Exemple :* si on souhaite connaître le nombre de fois où le classifieur a pris les battements normaux pour des battements inconnus, on regardera l'élément hors diagonal (2,3) dans la matrice de confusion.



Utilisation de la librairie *scikit learn* de Python (mais vous pouvez aussi les coder vous même !)

- réduction de dimension par ACP : fonction *PCA*
- méthode de classification classique k plus proches voisins (k-ppv) : fonction *KNeighborsClassifier*



*Exemple de classification par k-ppv.* L'échantillon de test (cercle vert) doit être classé soit dans la première classe des carrés bleus, soit dans la deuxième classe des triangles rouges. Si  $k = 3$  (cercle plein), il est assigné à la deuxième classe parce qu'il y a 2 triangles et seulement 1 carré à l'intérieur du cercle intérieur.

- **Parties IV** : Autres méthodes d'apprentissage supervisé : classification par **forêts aléatoires** et **réseaux de neurones**;
- **Partie V** : Réalisez **votre propre étude**, par exemple, en :
  - en augmentant les bases de test et/ou d'apprentissage
  - en proposant des variantes des méthodes proposées et/ou en testant d'autres méthodes de classification
  - en rajoutant davantage de classes (jusqu'à 2 supplémentaires : 4-Battelements ectopiques supraventriculaires, 5-Battelements de fusion)

Le tout en testant les approches (parties I à IV) et en interprétant les résultats via les mesures d'évaluation (matrice de confusion et pourcentage de données bien classées) et synthétiser votre étude dans un rapport (10 pages en pdf).