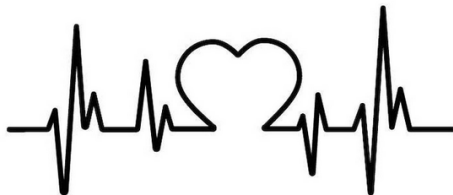


Projet :

Classification de battements cardiaques



Classification de battements cardiaques

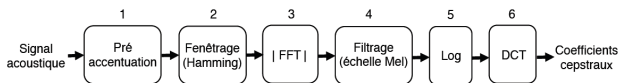
Les données ont été recueillies auprès de deux sources :

- (A) auprès du grand public via une application de smartphone,
- (B) dans le cadre d'un essai clinique dans des hôpitaux utilisant le stéthoscope numérique.

Les enregistrements de ces 2 sources étant de durées différentes, ils ont été ensuite transformés en MFCC pour extraire le contenu fréquentiel de ces données.

Transformation du signal temporel en MFCC (Mel Frequency Cepstral Coefficient)

Le cepstre présente l'avantage de permettre la séparation des contributions respectives de la source et du conduit vocal. Les MFCC s'obtiennent en utilisant, pour le calcul du spectre, une échelle fréquentielle non linéaire tenant compte de la perception auditive de la fréquence.



⇒ Utilisation de la librairie *Librosa*

Distance locale : distance euclidienne dans \mathbb{R}^p (ici $p = 20$)

$$d(X, Y) = \left(\sum_{k=1}^p (X_k - Y_k)^2 \right)^{\frac{1}{2}}$$

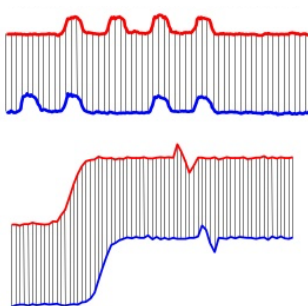
Organisation :

Ce projet en 6 séances se décomposent en 5 parties :

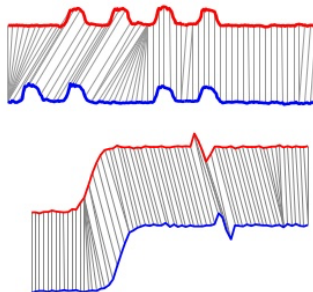
- *Partie I* : Système de reconnaissance d'activité physique avec la DTW
- *Partie II* : Réduction de dimension par ACP et classification par kppv
- *Partie III* : Classification par Forêts aléatoires
- *Partie IV* : Classification par réseaux de neurones
- *Partie V* : Votre étude

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 :

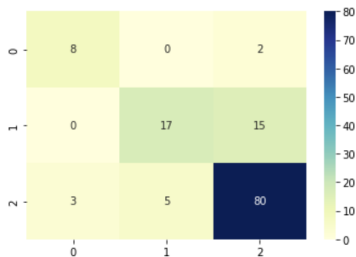
$$m = \arg_{1 \leq i \leq N} \min (\text{Score} (M, R_i))$$

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 .

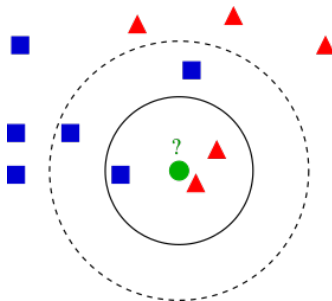
- Chaque ligne de la matrice de confusion représente la *classe réelle* tandis que chaque colonne représente une *classe prédite*.
- Les éléments diagonaux représenteront le nombre d'éléments bien classés. *Classification parfaite = matrice de confusion diagonale*
- Les éléments hors diagonaux représenteront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



Exemple : si on souhaite connaître le nombre de fois où le classifieur a pris des battements murmures (classe 1) pour des battements normaux (classe 2) 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 III & IV** : Classification par **forêts aléatoires** et **réseaux de neurones**;
- **Partie V** : Réalisez **votre propre étude**, par exemple, en :
 - en comparant les mesures des différents systèmes (set A et B)
 - en équilibrant les classes
 - en proposant des variantes des méthodes proposées et/ou en testant d'autres méthodes de classification

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 max en pdf).