

Mini Projet :

Segmentation d'images TEP par classification non supervisée

M1 IAFA-SECIL : Calcul Scientifique et Apprentissage Automatique

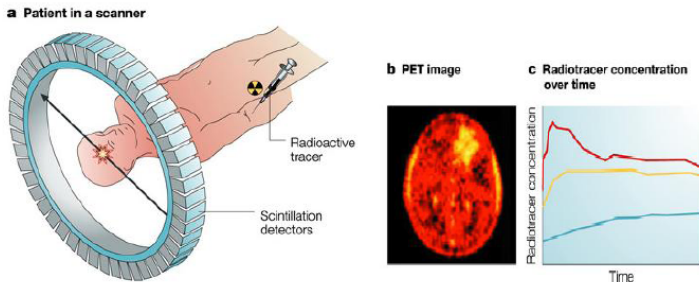


Figure: Principe de la Tomographie par Emission de Positons (TEP)

La Tomographie par Emission de Positons (ou TEP)

technique d'imagerie médicale fonctionnelle quantitative permettant de visualiser les activités du métabolisme. Les données TEP forment une séquence **3D + t** qui traduit l'évolution de la radioactivité dans le temps du volume correspondant au champ de vue de l'appareil de mesure.

Il est fourni les courbes temps activité (TAC) simulées d'une coupe saggitale et d'une coupe transverse.

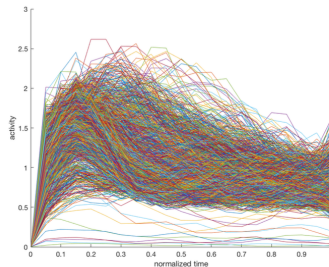
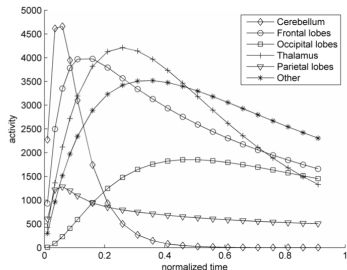


Figure: Courbe Temps Activité (TAC) : courbes théoriques et courbes simulées

Ici, l'information géométrique (position de chaque pixel dans l'image) n'est pas prise en compte.

Organisation :

Ce projet en 2 séances de TP encadrées se décomposent en 5 parties :

- *Partie I* : implémentation de la classification spectrale avec exemple jouet
- *Partie II* : application de la classification spectrale sur les profils temporels TACs
- *Partie III* : classification par Kmeans puis par réduction de dimension ACP+Kmeans
- *Partie IV* : évaluation des résultats
- *Partie V* : votre analyse

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 en pdf 5 pages max
- **Deadline** : le 28 octobre !

Algorithm 1 Algorithme de classification spectrale

Input : Ensemble des données S , Nombre de clusters k

1. Construction de la matrice affinité $A \in \mathbb{R}^{n \times n}$ définie par :

$$A_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2 / 2\sigma^2) & \text{si } i \neq j, \\ 0 & \text{sinon.} \end{cases}$$

2. Construction de la matrice normalisée $L = D^{-1}A$ où D matrice diagonale définie par :

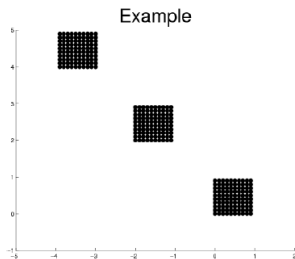
$$D_{i,i} = \sum_{j=1}^n A_{ij}.$$

3. Construction de la matrice $X = [x_1 x_2 \dots x_k] \in \mathbb{R}^{n \times k}$ formée à partir des k plus grands vecteurs propres x_i , $i = \{1, \dots, k\}$ de L .
4. Construction de la matrice Y formée en normalisant les lignes de X :

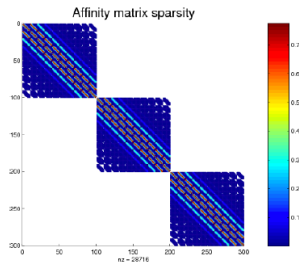
$$Y_{ij} = \frac{X_{ij}}{\left(\sum_j X_{ij}^2\right)^{1/2}}.$$

5. Traiter chaque ligne de Y comme un point de \mathbb{R}^k et les classer en k clusters via la méthode *K-means*.
 6. Assigner le point original x_i au cluster j si et seulement si la ligne i de la matrice Y est assignée au cluster j .
-

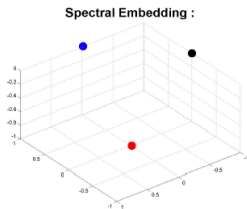
Figure: Algorithme de classification spectrale



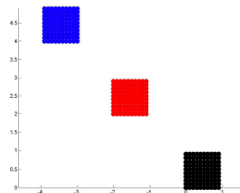
(a) Exemple pour $N = 300$ points



(b) Matrice affinité quasi bloc-diagonale (tronquée en dessous de 10^{-15})



(c) Lignes de Y



(d) 3 clusters séparés obtenus à la fin

FIGURE 1.1 – Illustration des étapes du clustering spectral

Travail demandé sur les parties I et II :

- Implémenter la méthode de classification spectrale
- Tester la fonction sur le jeu de données *ToyExample* et tester différentes valeurs de paramètres σ . Les résultats sont-ils similaires ?
- Tester la méthode de classification spectrale sur les courbes de temps activité TAC.

Travail demandé sur la partie III :

- Réaliser une classification avec le kmeans de scikit learn sur l'exemple jouet et les courbes TACs
- Tester aussi une réduction de dimension par ACP comme étape de prétraitement qui conserverait 95% de l'information puis classification avec le kmeans de scikit learn sur la représentation réduite des courbes TACs

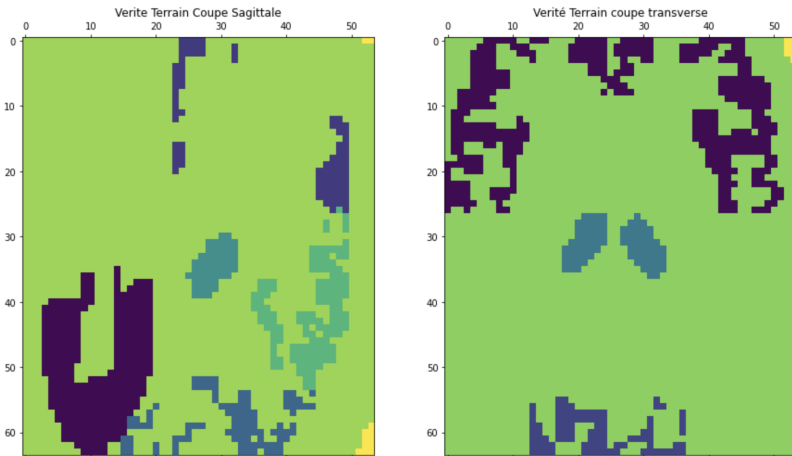


Figure: Images de la vérité terrain fournie pour les coupes saggitale et transverse

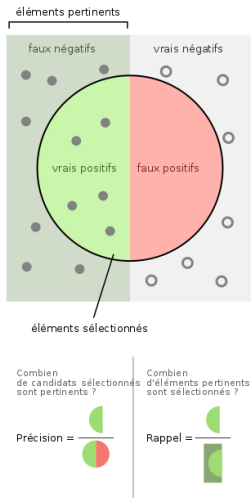


Figure: Illustration des mesures de précision et de rappel

Mesures d'évaluation :

Précision : évalue l'exactitude des prédictions positives

$$Precision = \frac{TP}{TP + FP}$$

Rappel : évalue le taux d'observations positives ayant été correctement détectées par le classifieur

$$Rappel = \frac{TP}{TP + FN}$$

Indice de Fowlkes-Mallows (FMI) mesure la similarité entre deux partitionnements. C'est une moyenne géométrique entre la précision et le rappel :

$$FMI = \frac{TP}{\sqrt{(TP + FP) * (TP + FN)}}$$

⇒ L'indice de Fowlkes-Mallows sera utilisé pour évaluer les résultats.

Travail demandé sur la partie V :

A partir des méthodes que vous avez implémentées, sur le notebook :

- Réaliser une analyse sur les résultats des méthodes de classification non supervisée : classification spectrale, kmeans et ACP+kmeans
- Tester pour différentes valeurs de paramètres σ et de classes
- Rédiger l'analyse effectuée et vos conclusions.

Le notebook sera à rendre sur moodle avec votre analyse rédigée.