

Algorithme des k-medoïdes

Véronique Tremblay

Différence par rapport aux k-moyennes

On n'utilise plus le centre, mais l'observations qui minimise les distances dans chaque groupe.

On utilise la fonction `pam` de la librairie `cluster` (Maechler et al. (2019)).

```
musique_pam <- pam(d, # La matrice de dissimilarité  
                  k = 6 # Le nombre de segments  
                  )
```

Dans les cas où le jeu de données est très volumineux, il pourrait être nécessaire de considérer d'utiliser la fonction `clara`.

Points forts

- Permet d'intégrer des variables nominales
- Robuste
- Permet de bien spécifier la matrice de distance

Points faibles

- Il faut connaître le nombre de groupes

- Similaire à l'algorithme des k-moyennes
- Permet d'utiliser des variables nominales

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2019. *Cluster: Cluster Analysis Basics and Extensions*.