

Méthodes hiérarchiques

Méthodes

Véronique Tremblay

Concept

Les méthodes hiérarchiques diffèrent selon leur façon de calculer la distance entre deux groupes.

Méthode du plus proche voisin (*single linkage*)

On utilise la dissemblance entre les observations les plus proches de chaque groupe.

Principales caractéristiques

- Tend à former un grand groupe avec plusieurs petits groupes satellites
- Permet de créer des groupes dont la forme est très irrégulière
- Est robuste aux données aberrantes

Méthode du voisin le plus distant (*complete linkage*)

On utilise la dissemblance entre les observations les plus éloignées de chaque groupe.

Principales caractéristiques

- Très sensible aux données aberrantes
- Tend à former des **groupes de tailles égales**

Méthode de la moyenne (*average linkage*)

On utilise la dissemblance entre deux groupes se définit comme étant la moyenne des dissemblance entre chaque individu des deux groupes.

Principales caractéristiques

- Groupes de faible variance et de même variance.

Principales caractéristiques

- Est optimale si les observations sont approximativement distribuées selon une loi normale multidimensionnelle de même matrice de variances-covariances
- Est sensible aux données aberrantes
- Tend à former des **groupes de même taille**

Généralisation de la plupart des approches en fonction de deux hyperparamètres.

Les propriétés dépendent des hyperparamètres et le choix des hyperparamètres n'est pas intuitif.

Méthode du centroïde

La distance entre deux groupes se définit comme étant la dissemblance entre les centroïdes des deux groupes.

Principales caractéristiques

- Robuste aux données aberrantes

La dissemblance entre deux groupes se définit comme étant la dissemblance médiane entre toutes les observations des deux groupes.

Principales caractéristiques

- Robuste aux données aberrantes

- Plusieurs méthodes
- La méthode à privilégier dépend du contexte
- Dans tous les cas:
 - Facilite le choix des groupes
 - Permet d'utiliser des variables nominales
 - Donne une hiérarchie