

# La classification hiérarchique

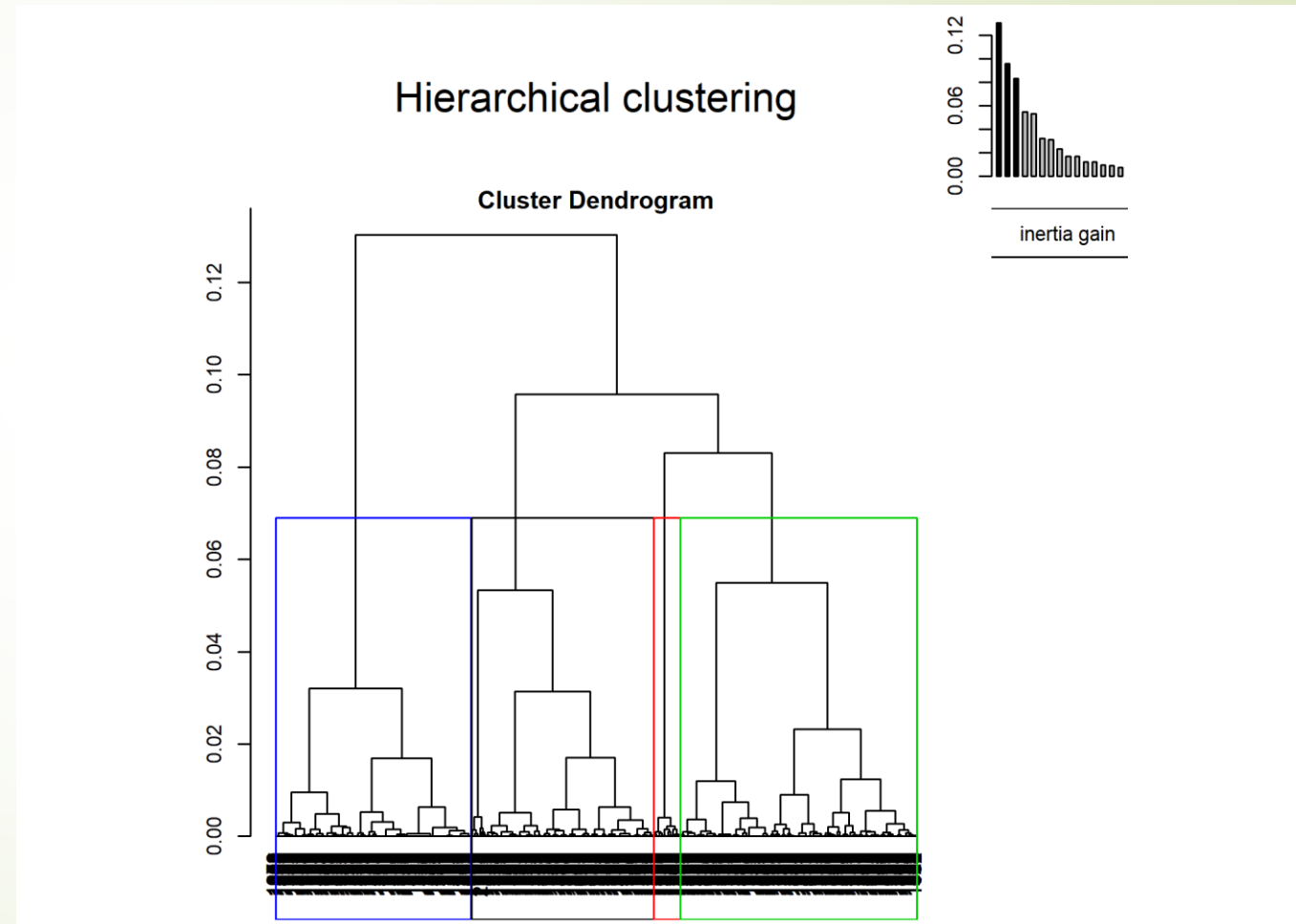
1

# Le clustering hiérarchique

Clusters emboîtés les uns dans les autres que l'on peut ainsi représenter dans le cadre d'une hiérarchie arborescente (**Dendrogramme**).

Plus on se situe bas dans l'arbre, plus les group se ressemblent.

On l'utilise généralement lorsqu'on ne connaît pas à l'avance le nombre de groups que l'on souhaite obtenir.



# Le clustering hiérarchique

Deux familles d'algorithmes :

- Les algorithmes ascendant (agglomératifs) : construisent les classes par agglomération successives des objets deux à deux.
- Les algorithmes descendants (divisifs) : réalise des dichotomies progressives de l'ensemble des objets

# Le clustering hiérarchique

Les algorithmes agglomératifs :

## 1 .Construction d'une matrice de distance entre les $m$ observations

- On regroupe chaque observation 2 à 2 avec celle dont elle est la plus proche
- Premier clustering avec  $m/2$  clusters de 2 observations.

## 2. Nouvelle matrice de distance qui résulte de l'agrégation précédente

- On cherche maintenant à regrouper chaque clusters 2 à 2 avec celui dont il est le plus proche.

## 3. Calcule des nouvelles distances

- On réitère l'opération jusqu'à n'avoir qu'un seul group.

# Le clustering hiérarchique

- Distance euclidienne standard:

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^n (x_{1j} - x_{2j})^2}$$

- Distance du Chi-2 (idéal pour comparer des proportions) -  $f$  étant la proportion:

$$d(x_1, x_2) = \sqrt{\frac{1}{f_n} \sum_{j=1}^n (f_{1j} - f_{2j})^2}$$

- Distance de Manhattan (utilisé pour minimiser l'influence des grands écarts):

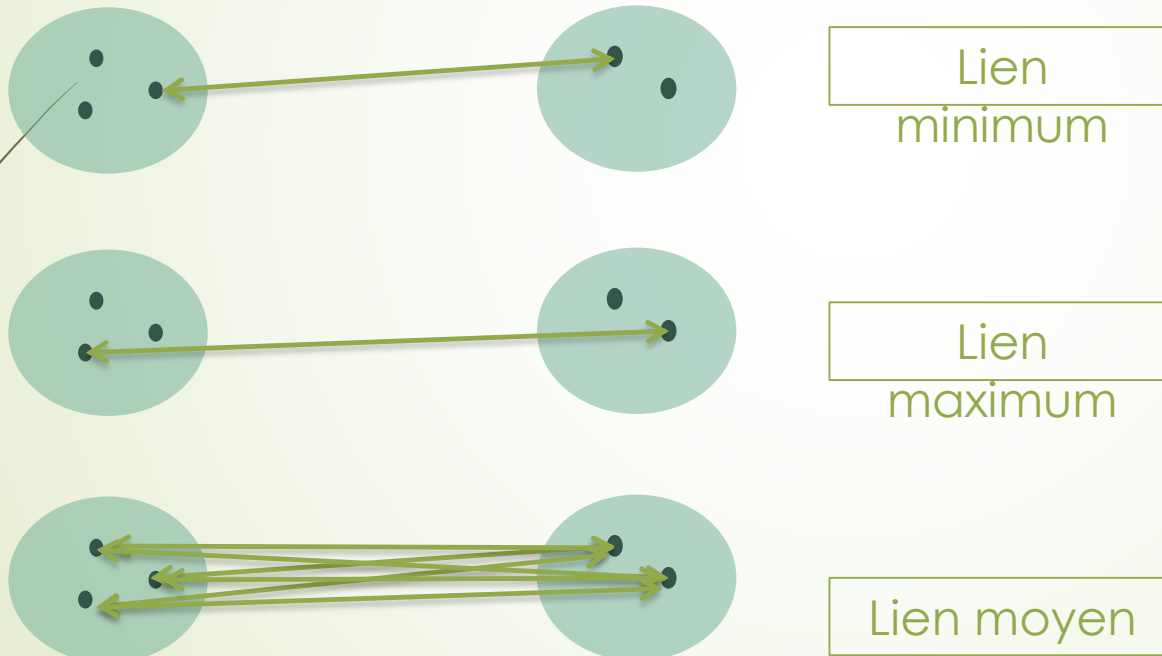
$$d(x_1, x_2) = \sum_{j=1}^n |x_{1j} - x_{2j}|$$

# Le clustering hiérarchique

6

► Le critère d'agrégation :

Lorsqu'on a des groupes agrégés, comment calculer la distance entre ces groupes ?



# Le clustering hiérarchique

- Autre critère : **Le critère de Ward** (la plus utilisé)

Celui-ci se base sur l'augmentation de l'inertie.

- **L'inertie intra-classe**, représente l'écart entre chaque point et le centre de gravité de la classe à laquelle il appartient.
- **L'inertie inter-classe**, représente l'écart entre chaque centre de gravité d'une classe et le centre de gravité général.

Utiliser le critère de Ward revient à agréger deux classes de façon à ce que l'augmentation de l'inertie intra-classe soit la plus petite possible.



# Le clustering hiérarchique

- Pour le découpage de dendrogramme on regarde :
  - L'allure général
  - Le nombre de clusters
  - La capacité à interpréter les clusters
- A noter que la longueur de chaque branche du dendrogramme est proportionnel à l'inertie inter-classe.



# Les approches mixtes

En mixant les approches hiérarchique et non hiérarchique on peut tirer partie des principaux avantages de ces méthodes :

- La capacité à analyser un grand nombre d'individus (classification non hiérarchique comme les kmeans notamment)
- Le choix d'un nombre de classe optimal (classification hiérarchique)

# Les approches mixtes

1. On utilise un clustering non hiérarchique pour obtenir rapidement un nombre de classes  $k$  supérieur au nombre de classes souhaité  $s$ .
2. On regroupe les  $k$  groupes de l'étape 1 avec un clustering hiérarchique cette fois ci.
  - Chaque groupe a pour poids la somme des poids qu'il représente).
  - Il convient d'utiliser le critère de « Ward » pour tenir compte des masses des classes.
3. On choisit le clustering final en choisissant les  $s$  groupes que l'on souhaite conserver.