

Activité 6 : catégorisation CAH

On souhaite effectuer la catégorisation automatique d'un ensemble de fromages.

L'objectif est d'identifier des groupes de fromages homogènes, partageant des caractéristiques similaires. Pour ce faire, la classification ascendante hiérarchique (CAH) sera utilisée.

Les données sont disponibles dans un fichier "fromage.txt". On possède 29 observations décrites par des variables quantitatives.

Travail à Faire

- 1- Importer les données à partir du fichier "fromage.txt".
- 2- Afficher la dimension de l'ensemble de données et en déduire le nombre des variables utilisées.
- 3- Afficher les noms de ces variables.
- 4- Déterminer l'intervalle de variation de chacune des variables.
- 5- Changer l'échelle des données pour les ramener dans l'intervalle [0-1].
- 6- Afficher le dendrogramme qui correspond à une classification hiérarchique ascendante.
 - a- Utiliser la distance euclidienne et tester plusieurs critères d'agrégation : "l'indice de Ward", "le lien minimum", "le lien maximum" et "le lien moyen".
 - b- Analyser les dendrogrammes obtenus en identifiant, pour chacun des critères, le meilleur choix du nombre de classes k.
- 7- Effectuer la catégorisation en utilisant la classification hiérarchique ascendante et visualiser les résultats obtenus et ce en utilisant :
 - Variables : seulement les 1^{er} et 2^{ème} attributs
 - Distance : euclidienne
 - Critères d'agrégation : "l'indice de Ward", "le lien maximum" et "le lien moyen".
 - Nombre de classes k : entre 2 et 5.

