

TD 2

L3 MIASHS - UBO

Exercice 1 : Convergence de l'algorithme K-means

Dans l'algorithme K-means, on note $S_j = \{C_1^{(j)}, \dots, C_k^{(j)}\}$ la partition en k classes obtenue à l'étape j et $G_i^{(j)}$, $i = 1, \dots, k$, les k centres de gravités des classes $C_i^{(j)}$.

1. Montrer que l'on a la relation suivante

$$I_{G_i^{(j)}}(C_i^{(j+1)}) = I_{G_i^{(j+1)}}(C_i^{(j+1)}) + \|G_i^{(j)} - G_i^{(j+1)}\|^2$$

En déduire que $I_{G_i^{(j+1)}}(C_i^{(j+1)}) \leq I_{G_i^{(j)}}(C_i^{(j+1)})$.

2. Montrer alors que $I_{intra}(S_{j+1}) \leq I_{intra}(S_j)$.
3. Qu'en déduit-on sur la convergence de l'algorithme K-means ?

Exercice 2 : Méthode de CAH "à la main"

On étudie trois caractères X_1, X_2 et X_3 sur cinq individus x_1, x_2, x_3, x_4 et x_5 . Les données sont les suivantes :

	X1	X2	X3
x1	1	2	3
x2	4	2	5
x3	4	3	7
x4	8	9	6
x5	4	2	3

Le but de cet exercice est d'effectuer une classification hiérarchique des individus avec la distance euclidienne et le critère du saut minimum.

1. Quelle est la matrice des données associée ?
2. Déterminer le centre de gravité G du nuage et calculer la matrice des distances entre les individus (arrondir au dixième les résultats).
3. Calculer l'inertie totale du nuage de points.
4. Faire une classification avec l'algorithme CAH avec le critère du saut minimum en mettant à jour la matrice des distances à chaque itération.
5. Tracer le dendrogramme associé.