

Classification Automatique

TP 1 : Distance, similarité et inerties

Un compte rendu est à rendre à la fin du TP. Il devra être constitué des programmes R et d'un rapport au format PDF. Les TP sont à réaliser en binôme. **La qualité de la rédaction (programmes et rapport) sera prise en compte.** Les programmes et le rapport devront porter les noms des binômes et être envoyés par mail à gwandji@polytech-lille.fr.

On considère 6 individus $\{w_1, w_2, w_3, w_4, w_5, w_6\}$ représentés par deux variables X_1 et X_2

$$\begin{pmatrix} 2.0 & 2 \\ 7.5 & 4 \\ 3.0 & 3 \\ 0.5 & 5 \\ 6.0 & 4 \\ 1.5 & 7 \end{pmatrix}$$

1. Construire les matrices des distances entre les individus en utilisant :
 - La distance de manhattan
 - La distance euclidienne.Comparer les résultats obtenus.
2. Ecrire une fonction qui calcule pour un ensemble de données les coordonnées de son barycentre. Appliquer la fonction à la partition $\{\{w_1, w_3\}; w_4; \{w_2, w_5\}; w_6\}$.
3. Ecrire une fonction qui calcule les distances entre un ensemble de points et son barycentre, et retourne la matrice des distances. On considère ici la distance euclidienne.
4. Calculer l'inertie totale de l'ensemble de données.
5. Calculer l'inertie interclasse, l'inertie intra classe et le pourcentage d'inertie expliqué. discuter.
6. Représenter graphiquement le nuage de point. Quelle segmentation proposeriez-vous ? Calculer le pourcentage d'inertie expliquée par cette nouvelle partition et comparer à celle de la question 2. Conclure.