

TD : Non-supervisée

1 Exercice 1 : CAH

Considérons la matrice de distance entre 4 observations :

| Obs | 1 | 2 | 3 | 4 |
|-----|-----|-----|------|------|
| 1 | 0 | 0.3 | 0.4 | 0.7 |
| 2 | 0.3 | 0 | 0.5 | 0.8 |
| 3 | 0.4 | 0.5 | 0 | 0.45 |
| 4 | 0.7 | 0.8 | 0.45 | 0 |

Par exemple, la distance entre la première et la deuxième observations est 0.3 et la distance entre la deuxième et la quatrième observations est 0.8.

1. Utiliser cette matrice de distance pour dessiner le dendrogramme résultant d'une classification ascendante hiérarchique avec la méthode du plus proche voisin. Indiquer sur le dendrogramme la hauteur à laquelle chaque fusion est faite.
2. Faire la même chose en utilisant la méthode du voisin le plus distant.
3. Quelle partition obtenez-vous si vous coupez ces deux arbres de sorte d'avoir 2 groupes ?
4. Il est possible d'échanger la position des feuilles de l'arbre sans changer l'interprétation du dendrogramme. Re-dessiner les dendrogramme obtenus précédemment en changeant la position de certaines feuilles, mais de sorte que les dendrogramme aient la même interprétation que ceux de base.

2 Exercice 2 : Indicateurs pour le choix des groupes

Soit a, b, c, d, e et f des observations qui ont respectivement les valeurs suivantes : 1, 2, 3, 4, 5, 6, 7. On les regroupe en deux. Les observations a, b et c sont dans le groupe 1 et les observations d, e et f sont dans le groupe 2.

1. Calculer l'indice de Silhouette de l'observation c . Interpréter le résultat.
2. Calculer le pseudo- R^2 de ce regroupement.

3 Exercice 3 : k -means

Dans cet exercice, on a 5 observations que l'on cherche à partitionner en $K = 2$ groupes. Le tableau suivant contient les observations :

| Obs | 1 | 2 | 3 | 4 | 5 |
|----------|----|------|-----|------|---|
| x_{i1} | -1 | -0.5 | 0 | 0.5 | 1 |
| x_{i2} | -1 | 0 | 0.5 | -0.5 | 1 |

1. Assigner aléatoirement les observations à chacun des deux groupes.
2. Calculer les centroïdes de chacun des groupes.
3. Calculer les distances entre les observations et les centroïdes. Calculer $W(C)$.
4. Ré-assigner les observations au groupe le plus proche.
5. Re-calculer les centroïdes de chacun des groupes.
6. Re-calculer les distances entre les observations et les centroïdes. Calculer $W(C)$. Les observations changent-elles de groupe ?