

Réduction de la dimension

1 Fléau de la dimension, contraste

A partir du jeu de données `mat.txt`, qui est une matrice de taille 1000×1000 , calculez pour chaque ligne la norme des d premières colonnes (faites varier d sur une grille) et faites pour les d que vous avez choisi le rapport du max de la norme sur le min de la norme. Dessinez ce contraste en fonction de votre grille.

2 Classification du jeu de données `donclassif.txt`

Ce jeu de données est le jeu de données illustrant le cours.

1. Importer les données
2. Effectuer une classification hiérarchique
3. Effectuer une classification avec k -means en faisant varier k
4. Effectuer une classification avec DBSCAN en faisant varier `DistMin` et ε .
5. Comparer les résultats obtenus
6. Idem avec `donclassif2.txt`, pour ce jeu de données là, il est impossible de faire une CAH directement, il faut donc faire une k -means avec beaucoup de groupes et enchaîner par une CAH.

3 Classification du jeu de données `temperature.txt`

Ce jeu de données donne les températures de 15 villes françaises pendant 12 mois.

1. Importer les données
2. Effectuer une classification hiérarchique
3. Effectuer une classification avec k -means en faisant varier k
4. Effectuer une classification avec DBSCAN en faisant varier `DistMin` et ε .
5. Comparer les résultats obtenus
6. Pour aller plus loin, dessiner les villes sur une carte en mettant en couleurs l'appartenance au groupe.

4 ACP

Ce jeu de données donnent les résultats aux épreuves de décathlon lors de deux compétitions : les JO et au Décastar pour un certain nombre de décathlons de haut niveau. Les variables sont donc les performances aux 10 épreuves, le classement obtenu, le nombre de points et le lieu de la compétition.

Commentez l'ACP de ce jeu de données.

Rappel : les étapes sont (toujours les mêmes) :

1. Importer les données
2. Sélectionner les variables à utiliser
3. Sélectionner le nombre d'axes
4. Commenter la représentation des variables (1-2 ; 3-4 etc.).
5. Commenter la représentation des individus (si cela a un intérêt) avec l'aide des \cos^2 .
6. Représenter les variables supplémentaires (s'il y en a).