

Introduction

Dans ce projet il est demandé d'effectuer une classification sur les données de consommation d'électricité de 2914 logements irlandais, et ce afin d'obtenir des classes homogènes, séparant les logements en plusieurs profils, selon leurs consommation.

Les différents algorithmes utilisés, vus durant le cours d'apprentissage non supervisé, sont :

- La classification hiérarchique ascendante(CAH).
- L'algorithme des centres mobiles (K-means).
- L'algorithme Self Organising Map (SOM).
- La classification spectrale.

Les algorithmes sont directement utilisés sur les données fournies, car celles-ci ont déjà été normalisées.

Dans les prochains chapitres, ce rapport décrira l'utilisation des différents algorithmes sous R sur le jeu de données, puis discutera les résultats, premièrement sur l'ensemble des données, puis dans un second temps uniquement sur les données représentant les heures des jours ouvrés. Ce document se termine par une brève conclusion.

Partitionnement et analyse sur l'ensemble des données

Algorithmes de classification

Classification Hiérarchique Ascendante CAH

Les algorithmes de classification hiérarchique ascendante sont des algorithmes qui forment des clusters à chaque itération en combinant plusieurs clusters, se basant sur la notion de distance entre ceux-ci.

À l'initialisation de l'algorithme, chaque instance représente un cluster. À la fin de l'algorithme, toutes les instances se trouvent dans un seul et même cluster.

Ici, la distance utilisée est la distance euclidienne. Pour commencer les distances sont calculées deux à deux entre instances et sont sauvegardées dans une matrice symétrique dite "Matrice de distance".

La méthode de combinaison des clusters utilisée ici est appelée méthode de Ward, cette méthode permet, à chaque étape, d'augmenter un minimum l'inertie intra-classe, et ce jusqu'à ce que celle-ci atteigne la valeur de l'inertie totale.

Continuer ici dendrogramme séparation des classes ...

Algorithme des Centres Mobiles K-means

Algorithme Self Organising Map

Classification Spectrale

Représentations et analyses des résultats

Partitionnement et analyse sur les données hors week-end

Algorithmes de classification

Représentations et analyses des résultats

Conclusion