

Model-based learning

TP3 : clustering

Dans ce TP nous allons travailler sur les données `Velib` disponibles dans le package `funFEM`. Ces données représentent les courbes de charge des stations Velib à Paris sur environ une semaine. On dispose d'une donnée toutes les heures.

1. Installer le package `funFEM` et récupérer les données `Velib`.
2. Représenter graphiquement quelques courbes de charges pour différentes stations. Repérez-vous des profils particuliers ?
3. A l'aide d'une analyse en composante principale (ACP), représenter les données et tenter d'en donner une interprétation. Il pourra être intéressant de ne travailler ici que sur une partie des variables (par exemple en choisissant de ne regarder que quelques heures types chaque jour) de sorte à faciliter la visualisation des résultats. Quelle interprétation pouvez-vous tirer de cette analyse ?
4. Réaliser un clustering de ces données à l'aide de modèles de mélanges gaussiens. Choisissez le modèle et le nombre de clusters de façon adéquate. Représenter graphiquement les résultats (par exemple en représentant les moyennes de chaque groupe). Interprétez les résultats.
5. Le package `ggmap` permet de réaliser des *représentations géographiques* sous R. Utilisez-le pour représenter les clusters obtenus. Pouvez-vous interpréter ces résultats ?