

## Projet d'apprentissage non supervisé

à rendre pour le 31/10/2019  
remise possible jusqu'au 07/11/2019

### 1 Données

Le jeu de données (<https://archive.ics.uci.edu/ml/machine-learning-databases/00482>) analysé dans ce projet est relatif au taux de remplissage de parcs de stationnement d'automobiles à Birmingham. Il est caractérisé par 4 variables :

- `SystemCodeNumber` : identifiant du parking,
- `Capacity` : capacité du parking,
- `Occupancy` : nombre de places disponibles dans le parking au moment de la mesure,
- `LastUpdated` : date et heure de mesure du nombre de places disponibles.

L'objectif du projet est d'exploiter la classification non supervisée afin de regrouper en classes homogènes les parking ainsi que leurs taux d'occupation. Une telle classification pourra par exemple servir à améliorer les prévisions sur ce taux d'occupation, en les réalisant classe par classe.

### 2 Travail à effectuer

- 1) Préparation des données : après avoir mis en forme les données, repérer les valeurs aberrantes, erronées ou manquantes. Comblent celles-ci quand cela vous paraît possible.
- 2) Statistiques descriptives : décrire les données à l'aide de statistiques et de graphiques élémentaires que vous jugerez pertinents pour ce problème.

- 3) Analyse du comportement hebdomadaire : réaliser une classification de l'ensemble des séries temporelles hebdomadaires décrivant le taux de charge des parkings. Utiliser pour cela au moins deux méthodes étudiées en cours. Donner une interprétation des résultats obtenus.

*Indication : Soit  $P$  le nombre de parkings étudiés,  $S$  le nombre de semaines et  $T$  le nombre de mesures disponibles par semaine. Le jeu de données à classer ici comprendra  $P \times S$  lignes et  $T$  colonnes.*

- 4) Analyse des parkings : regrouper les parkings en classes homogènes. Utiliser pour cela au moins deux méthodes étudiées en cours. Donner une interprétation des résultats obtenus.

*Indication : Pour cette question, la taille de la série décrivant chaque parking peut être réduite, par exemple en considérant une mesure (moyenne) par jour ; mais cette réduction n'est pas obligatoire.*

### 3 Documents à rendre

Deux fichiers : un rapport en pdf n'excédant pas 15 pages et un fichier contenant le code R associé.

### Références

- [1] Stolfi D.H., Alba E., Yao X. (2017). Predicting car park occupancy rates in smart cities. In : Alba E., Chicano F., Luque G. (eds) Smart Cities. Smart-CT 2017. Lecture Notes in Computer Science, vol 10268. Springer.
- [2] Birmingham City Council. <https://data.birmingham.gov.uk/dataset/birmingham-parking>