

TME 2 - Estimation de densité

Données : API Google Places, Points d'intérêt de Paris

Télécharger l'archive sur le site de l'ue. Elle contient un dictionnaire avec des informations sur différents types de points d'intérêt (POI) de Paris¹. Chaque clé du dictionnaire correspond à un type de POI (restaurant, distributeur d'argent, ...) et la valeur associée est un dictionnaire avec l'identifiant google map comme clé, et les informations sur le POI : les coordonnées GPS, sa notation, son nom, ses types, le niveau des prix.

Le code fourni permet de visualiser les données et charger la matrice des coordonnées GPS.

Expérimentations

L'objectif de ce TME est de prendre en main les algorithmes d'estimation de densité. Choisissez un type de POI à traiter, l'objectif est d'estimer la loi de densité $p(x, y)$ géographique des POIs sur Paris (x, y sont les coordonnées GPS d'un lieu). Vous étudierez les deux algorithmes suivants :

- Méthode des histogrammes.
- Méthode à noyaux (Parzen pour commencer puis le noyaux de votre choix - gaussien par exemple).

Pour chacune des expériences, vous visualiserez les résultats obtenus en faisant une discrétisation de l'espace et en évaluant en chaque case la densité donnée par vos modèles (la méthode par histogramme vous donne directement la discrétisation à employer). Le code pour l'affichage vous est fourni.

Expérimentez selon différents réglages. Réfléchissez en particulier aux questions suivantes :

- Que se passe-t-il pour une faible/forte discrétisation pour la méthode des histogrammes ?
- Quel est le rôle des paramètres des méthodes à noyaux ?
- Comment choisir de manière automatique les meilleurs paramètres ?
- La question reliée : comment estimer la qualité de votre modèle ?

Vous pouvez par la suite essayer de comparer la densité des différents type de POI. Vous pouvez également vous intéresser à un problème de classification qui consiste à prédire la note d'un POI en fonction de son emplacement. Dans ce dernier contexte de classification supervisée, vous pouvez mettre en œuvre l'estimateur de Nadaraya-Watson et les k -plus proches voisins.

1. Le script pour récupérer d'autres POI est fourni, il s'agit de `gmap-poi.py`, il vous faut pour l'utiliser une clé API de développement google, gratuite.