

Master Modélisation Statistique 2

TP APPRENTISSAGE STATISTIQUE : MODÉLISATION NON-PARAMÉTRIQUE PAR DES FONCTIONS B-SPLINES UNI-VARIÉES ET BI-VARIÉES

Ce projet doit être réalisé en trinôme, avant le 4 décembre. La taille maximale du projet (sans le bonus et sans les annexes) est de 5 pages. Le bonus ne doit pas dépasser 2-3 pages.

Le jeu de données sur lequel vous allez travailler concerne les prix de vente des appartements de Varsovie (Pologne) :

```
> require(HRW)
> data(WarsawApts)
```

Ce jeu de données contient les variables suivantes :

```
Surface: surface en mètres carrés
district: variable facteur à quatre modalités (districts ou cartiers):
  Mokotow, Srodmiescie, Wola and Zoliborz.
n.rooms: nombre de chambres
floor: l'étage
construction.date: année de construction
areaPerMzloty: prix au mètre carré
```

La variable à prédire est `areaPerMzloty` et vous devez utiliser les méthodes non-paramétriques vues en cours. Vous pouvez commencer par prédire cette variable en utilisant une seule variable explicative (c'est à vous de choisir la variable explicative et justifier qu'un modèle non-paramétrique est conseillé) et compliquer ensuite votre modèle en utilisant plusieurs variables explicatives. Vous ne devez pas implémenter toutes les méthodes vues en cours, mais de choisir celle qui vous semble adaptée. Vous pouvez choisir les paramètres de lissage par validation croisée par blocks.

Bonus : modélisation non-paramétrique avec des variables explicatives quantitatives et qualitatives

Comme vous pouvez le remarquer, le jeu des données contient une variable facteur qui peut être utilisée dans le modèle (de façon linéaire ou non-linéaire) en même temps que les autres variables quantitatives. Nous n'avons pas implémenté ce type de modèle non-paramétrique qui combine variable quantitative et qualitative dans le tp mais vous avez le descriptif dans le TP 2 de la mise en oeuvre avec la fonction `gam` de ce type de modèles. Grosso-modo, l'interprétation de ce type de modélisation est qu'on veut avoir un modèle (paramétrique ou non-paramétrique) différent par modalité du facteur.