

8. RÉGRESSION PAR RÉGRESSOGRAMME

RÉGRESSION NON PARAMÉTRIQUE : RÉGRESSOGRAMME

Le régressogramme, et en quelque sorte l'équivalent de l'histogramme vu pour l'estimation non paramétrique d'une densité de probabilité.

1. Partitionner l'espace $[m, m + l]$ contenant les observations de la variable X en b intervalles de taille ν .
2. Pour tout intervalle $[m + (i - 1)\nu, m + i\nu], i = 1, \dots, b$, procéder au comptage des x_j :

$$\forall i = 1, \dots, b, C_i := \sum_{j=1}^N \mathbb{I}_{[m+(i-1)\nu, m+i\nu]}(x_j)$$

3. Enfin, pour chaque $x \in [m + (i - 1)\nu, m + i\nu]$, la i -ème boîte, et pour chaque $i = 1, \dots, b$, nous estimons la fonction de régression en prenant la moyenne des y_i correspondantes à la classe considérée :

$$\hat{g}_\nu^{\text{Reg}}(x) = \frac{\sum_{j=1}^N \mathbb{I}_{[m+(i-1)\nu, m+i\nu]}(x_j) y_j}{C_i}$$

RÉGRESSION NON PARAMÉTRIQUE : RÉGRESSOGRAMME

Les mêmes observations faites à propos des limites de cette approche s'appliquent ici également : il s'agit d'une approximation constante par morceaux. Pour l'obtenir, nous avons du choisir a priori certains paramètres :

- La valeur (ou coordonnée) m
- Le nombre total d'intervalles (ou boîtes) b
- La longueur (ou volume) de chaque intervalle/boîte ν

9. RÉGRESSION PAR NOYAUX

RÉGRESSION NON PARAMÉTRIQUE : NOYAUX

Comme vu pour le cas de l'estimation de la densité, dans ce cas également une manière naturelle pour rendre l'estimation plus lisse est celle d'utiliser des noyaux K (un paramètre de lissage ν est à définir ici également).

$$\hat{g}_{\nu}^K(x) = \frac{\sum_{j=1}^N K\left(\frac{x_j - x}{\nu}\right) y_j}{\sum_{j=1}^N K\left(\frac{x_j - x}{\nu}\right)}$$

RÉGRESSION NON PARAMÉTRIQUE : NOYAUX

Les mêmes noyaux peuvent être reproposés ici :

- Le noyau gaussien : $K(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
- Le noyau d'Epanechnikov : $K(z) := \frac{3}{4} (1 - z^2) \mathbb{I}_{[-1,1]}(z)$
- Le noyau triangulaire : $K(z) := (1 - |z|) \mathbb{I}_{[-1,1]}(z)$
- Le noyau uniforme : $K(z) := \frac{1}{2} \mathbb{I}_{[-1,1]}(z)$

Télécharger le fichier TP5_Regression_partie2.ipynb : ibalelli.github.io → Teaching → Modélisation statistique avancée

10. PRÉDICTION NON PARAMÉTRIQUE

PRÉDICTION

À quoi pouvez vous vous attendre lorsque vous souhaitez effectuer une prédiction pour un nouveau point x ? Si x est à l'intérieur de l'intervalle d'entraînement ? Et si x est à l'extérieur ?

PRÉDICTION

