

1 Introduction

En pratique, les modèles linéaires ne sont pas toujours bien adaptés, on a alors la possibilité d'utiliser d'autres alternatives, à savoir : les polynômes, les splines, les modèles additifs généralisés, les fonctions en escalier, etc.

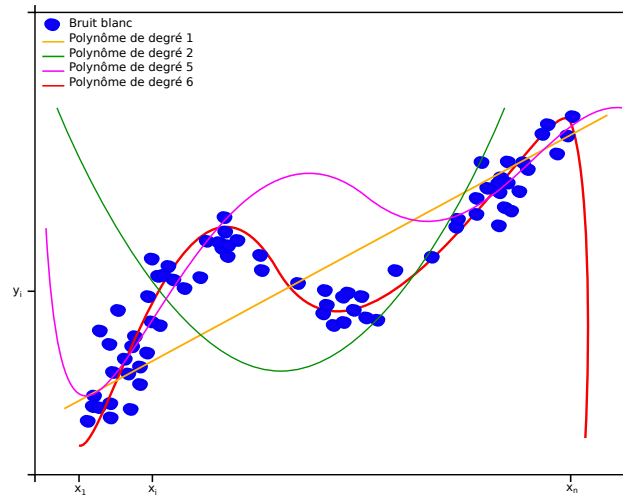


FIGURE 1 – Les limites du modèle linéaire

2 Modèle polynomial

On suppose que l'on dispose de l'observation de n variables aléatoires réelles $y_i, i = 1, \dots, n$ indépendantes, alors un modèle polynomial de degré D s'écrit sous la forme :

$$y_i = \beta_0^* + \sum_{j=1}^D \beta_j^* x_i^j + \varepsilon_i, \quad i = 1, \dots, n$$

Où les $x_i, i = 1, \dots, n$ sont les observations de la variable explicative X et on suppose que les variables aléatoires ε_i sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

L'écriture matricielle du modèle :

$$y = X\beta^* + \varepsilon$$

avec $X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & \cdots & x_2^D \\ 1 & x_3 & x_3^2 & \cdots & x_3^D \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^D \end{pmatrix}$ qui est une matrice de **Vandermonde** de n lignes et $D+1$ colonnes,

où l'on note $X_{i,j} = x_i^{j-1}$ et $\beta^* = (\beta_0^*, \dots, \beta_D^*)^T \in \mathbb{R}^{D+1}$.

Remarque. On a $D+1$ paramètres pour un polynôme de de degré D et ils sont estimés par la méthode des moindres carrés

2.1 Choix du degré

On peut utiliser la Validation Croisée pour choisir le degré du polynôme ou en d'autres critères de sélection de type AIC, BIC, etc.

2.2 Avantages et inconvénients d'un modèle polynomiale

Les avantages

- utile en estimation non-paramétrique,
- flexibilité pour de faibles degrés.

Les inconvénients et limites

- les polynômes ne sont pas localisés (fonctions globales, non localisées),
- le nombre de paramètres à estimer augmente vite avec le degré (et la dimension),
- effet de bord et oscillations (mauvaise prédiction en dehors de la gamme de valeurs observées).

Exemple. On prend $n = 2$ et $D = 2$, considérons $x_i \in \mathbb{R}^2 : x_i = \begin{pmatrix} x_{i,1} \\ x_{i,2} \end{pmatrix}$

et

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,1}^2 & x_{1,2}^2 & x_{1,1}x_{1,2} \\ 1 & x_{2,1} & x_{2,2} & x_{2,1}^2 & x_{2,2}^2 & x_{2,1}x_{2,2} \end{pmatrix}.$$

Un polynôme de degré 2 requiert de fixer les coefficients liés aux variables : $[x_{i,1}, x_{1,2}, x_{i,1}^2, x_{i,2}^2, x_{i,1}x_{i,2}]$

3 Splines (Cerces)

Cette méthode non paramétrique a été développée dans les années 1950/60's notamment par Pierre Bézier.

Définition. Soit f une fonction polynomiale par morceaux sur un intervalle $[a, b]$, $f : [a, b] \rightarrow \mathbb{R}$, composée de n sous intervalles $[x_{i-1}, x_i]$ avec $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$. La restriction de f sur chaque intervalle $[x_{i-1}, x_i]$ est un polynôme $P_i : [x_{i-1}, x_i] \rightarrow \mathbb{R}$, ainsi

$$f(x) = \begin{cases} P_1(x) & \text{si } x \in [x_0, x_1] \\ P_2(x) & \text{si } x \in [x_1, x_2] \\ \vdots & \\ P_n(x) & \text{si } x \in [x_{n-1}, x_n] \end{cases}$$

Le plus haut degré des polynômes P_i est appelé l'ordre du spline f , et les x_i sont appelés les noeuds

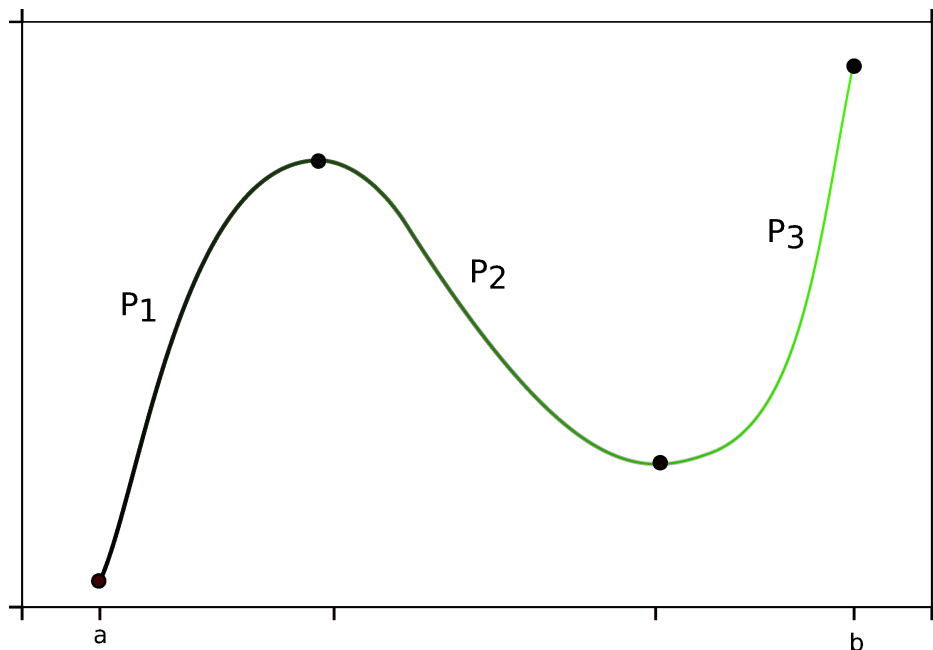


FIGURE 2 – Exemple d'un spline de degré 3

Remarque. Les splines les plus populaires sont les splines (cubiques) d'ordre 3

Remarque. On privilégie des splines lisses : C_0, C_1, C_2 , etc.

3.1 Cadre d'utilisation

Les splines sont utilisés dans des problèmes d'interpolation et de lissage pour représenter numériquement des contours complexes, dans des domaines tels que :

- statistique,
- computer vision,
- analyse numérique.

3.2 Estimation par splines

Pour ajuster des splines quand on observe des points $(x_i, y_i) \in \mathbb{R} \times \mathbb{R}$ pour $i = 1, \dots, n$ on cherche le spline avec une courbure minimum, *i.e.*, résoudre :

$$\hat{f} = SP_\lambda(y) \in \arg \min_{f \text{ est un spline}} \left(\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 \right) + \lambda \int_a^b |f''(t)|^2 dt .$$

Remarque. Quand

$$\lambda \rightarrow \infty, \text{ on a nécessairement } \int_a^b |f''(t)|^2 dt = 0 \Rightarrow f'' = 0 \Rightarrow f' = \text{cte}$$

alors f est un polynôme de degré $D = 1$ et donc affine par morceaux, sur chaque intervalle, et peut donc interpoler tous points.

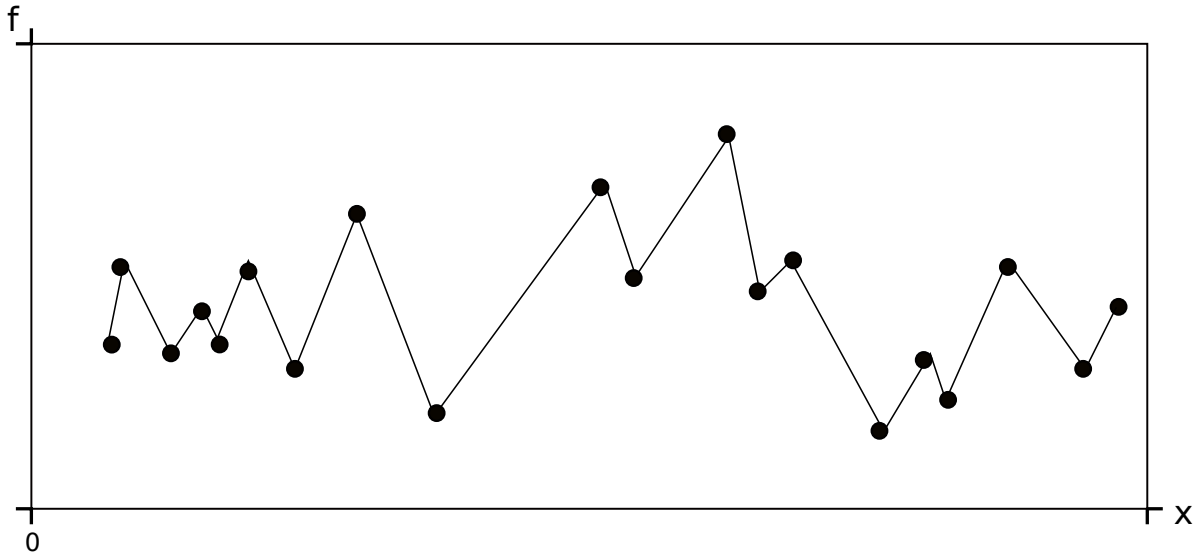


FIGURE 3 – Polynôme d’interpolation de Lagrange (très approché!)

La solution : la solution est atteinte pour un spline cubique (d’ordre 3), et peut être obtenue par un moindre carré régularisé, pour une certaine matrice $\Omega \in \mathbb{R}^{n \times n}$:

$$\arg \min_{g \in \mathbb{R}^n} \|y - g\|^2 + \lambda g^T \Omega g .$$

Note : avec cette régularisation les splines ont pour nœuds les x_i .

Remarque. Numériquement, on peut résoudre ce problème comme un problème d’optimisation quadratique.

4 Modèles additifs généralisés (GAM)

Avec les des fonctions réelles, *i.e.*, $f_j : \mathbb{R} \rightarrow \mathbb{R}$, le modèle s’écrit

$$y_i = \sum_{j=1}^p f_j(x_{i,j}) + \varepsilon_i \text{ ou } y = \sum_{j=1}^p f_j(\mathbf{x}_j) + \varepsilon$$

avec

$$f_j(\mathbf{x}_j) = \begin{pmatrix} f_j(x_{1,j}) \\ \vdots \\ f_j(x_{n,j}) \end{pmatrix} \text{ et } \mathbf{x}_j = \begin{pmatrix} x_{1,j} \\ \vdots \\ x_{n,j} \end{pmatrix}$$

et où les ε_i sont i.i.d. centrés de variance σ^2 ,

Remarque. En optimisation l’intérêt est le suivant :

$$\arg \min_{\mathbf{x}_1, \dots, \mathbf{x}_p} \left(\sum_{j=1}^p f(\mathbf{x}_j) \right) = \left(\arg \min_{\mathbf{x}_1} f(\mathbf{x}_1), \dots, \arg \min_{\mathbf{x}_p} f(\mathbf{x}_p) \right)$$

Il remplace un "gros" problème d'optimisation en p petits "problèmes" (de dimension 1)

Remarque. potentiellement un des f_j encode la variable constante.

4.1 Algorithme : Rétro-ajustement d'un modèle additif (Backfitting)

Entrées : $(\mathbf{x}_i)_{i=1,\dots,n}, (\mathbf{y}_i)_{i=1,\dots,n}, (\lambda_j)_{j=1,\dots,p}$

Initialisation : $f_1 = 0, \dots, f_p = 0$ et $\mathbf{r} = \mathbf{y}$ (résidu)

tant que la convergence n'est pas atteinte **faire**

pour $j = 1, \dots, p$ **faire**

$\mathbf{r} = \mathbf{r} + f_j(\mathbf{x}_j)$

$f_j = SP_{\lambda_j}(\mathbf{r})$

$\mathbf{r} = \mathbf{r} - f_j(\mathbf{x}_j)$

retourner $(f_1, \dots, f_p) = f$

4.2 Avantages et inconvénients des GAM

Avantages

- peut modéliser des effets non-linéaires automatiquement,
- interprétation possible : fonctions 1D (visualisation possible),
- peut s'étendre au cas d'interactions d'ordre plus élevées (p petit).

Inconvénients

- Calibrage difficile : au moins un paramètre par variable, en général, on suppose que $\lambda_1 = \lambda_2 = \dots = \lambda_p$ pour simplifier ce choix ;
- Critère d'arrêt n'est pas si simple (non-convexe).

Remarque. Cet algorithme s'apparente à l'algorithme d'optimisation "descente par coordonnée" qui cherche à résoudre le genre de problème suivant :

$$\min_{x_1, \dots, x_p} f(x_1, \dots, x_p).$$

Algorithme du descente par coordonnée :

Initialisation : $x = 0$

tant que la convergence n'est pas atteinte **faire**

pour $j = 1, \dots, p$ **faire**

$x_j \leftarrow \arg \min_{x_j} f(x_1, \dots, x_{j-1}, x_j, \dots, x_p).$

Références

[SAL18] J. SALMON and N. VARZELEN, Méthodes non-linéaires : Polynômes, modèles par morceaux, splines et GAM, Université de Montpellier, 2018