ML 01 Méthode des moindres carrés

1 Espace de fonctions

Soit un espace vectoriel composé de fonctions. Une base de cet espace est un ensemble de fonctions $(f_1, f_2, \dots f_N)$ tel que toute fonction de l'espace s'exprime comme combinaison linéaire des fonctions de base.

$$f(\mathbf{x}) = a_1 f_1(\mathbf{x}) + a_2 f_2(\mathbf{x}) + \cdots + a_N f_N(\mathbf{x})$$

Pour un jeu de données $\left\{\mathbf{x^{(k)}}, y^{(k)}\right\}_{k=1}^{N}$ de taille N, les coefficients a_i sont solution d'un système linéaire.

$$\begin{pmatrix} f_1(\mathbf{x^{(1)}}) & f_2(\mathbf{x^{(1)}}) & \dots & f_N(\mathbf{x^{(1)}}) \\ f_1(\mathbf{x^{(2)}}) & f_2(\mathbf{x^{(2)}}) & \dots & f_N(\mathbf{x^{(2)}}) \\ \dots & \dots & \dots & \dots \\ f_1(\mathbf{x^{(N)}}) & f_2(\mathbf{x^{(N)}}) & \dots & f_N(\mathbf{x^{(N)}}) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_N \end{pmatrix} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N)} \end{pmatrix}$$

Nous notons ce système linéaire $\mathbf{Ax} = \mathbf{b}$.

2 Expression matricielle

Le système linéaire $\mathbf{Ax} = \mathbf{b}$ avec $\mathbf{A} \in \mathbb{R}^{M \times N}$ n'a pas de solution quand le nombre d'observations dépasse le nombre de fonctions de base (c'est-à-dire, $M > N$). Une approche possible est alors de chercher une approximation $\mathbf{Ax} \approx \mathbf{b}$ qui minimise la somme des carrés des erreurs : $\|\mathbf{Ax} - \mathbf{b}\|_2^2$.

$$\begin{aligned}
& \|\mathbf{Ax} - \mathbf{b}\|_2^2 \\
= & \{\|\mathbf{x}\|_2 = \sqrt{\mathbf{x} \cdot \mathbf{x}}\} \\
& (\mathbf{Ax} - \mathbf{b}) \cdot (\mathbf{Ax} - \mathbf{b}) \\
= & \{\text{Par définition du produit scalaire euclidien}\} \\
& (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}) \\
= & \{\text{propriété de la transposition}\} \\
& \left(\mathbf{x}^T \mathbf{A}^T - \mathbf{b}^T\right) (\mathbf{Ax} - \mathbf{b}) \\
= & \{\text{multiplication}\} \\
& \mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - \mathbf{x}^T \mathbf{A}^T \mathbf{b} - \mathbf{b}^T \mathbf{Ax} + \mathbf{b}^T \mathbf{b} \\
= & \{\mathbf{b}^T \mathbf{Ax} \text{ étant une valeur scalaire, } \mathbf{b}^T \mathbf{Ax} = \left(\mathbf{b}^T \mathbf{Ax}\right)^T = \mathbf{x}^T \mathbf{A}^T \mathbf{b}\} \\
& \mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - 2\mathbf{x}^T \mathbf{A}^T \mathbf{b} + \mathbf{b}^T \mathbf{b}
\end{aligned}$$

Cette dernière expression quadratique en $\mathbf{x}$ correspond à une surface convexe. Donc son minimum peut être calculé en annulant sa dérivée (penser à une courbe $y = a + bx + cx^2$ dont l'unique extremum est atteint lorsque la pente est nulle).

$$\begin{aligned}
& \mathbf{0} = 2\mathbf{A}^T \mathbf{Ax} - 2\mathbf{A}^T \mathbf{b} \\
= & \\
& \mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}
\end{aligned}$$

Ainsi, quand $M > N$, la solution approximée $\mathbf{x}$, telle que $\mathbf{Ax} \approx \mathbf{b}$ par minimisation de la somme des carrés des erreurs, est la solution du système linéaire suivant où $\mathbf{A}^T\mathbf{A}$ est appelée la matrice de Gram.

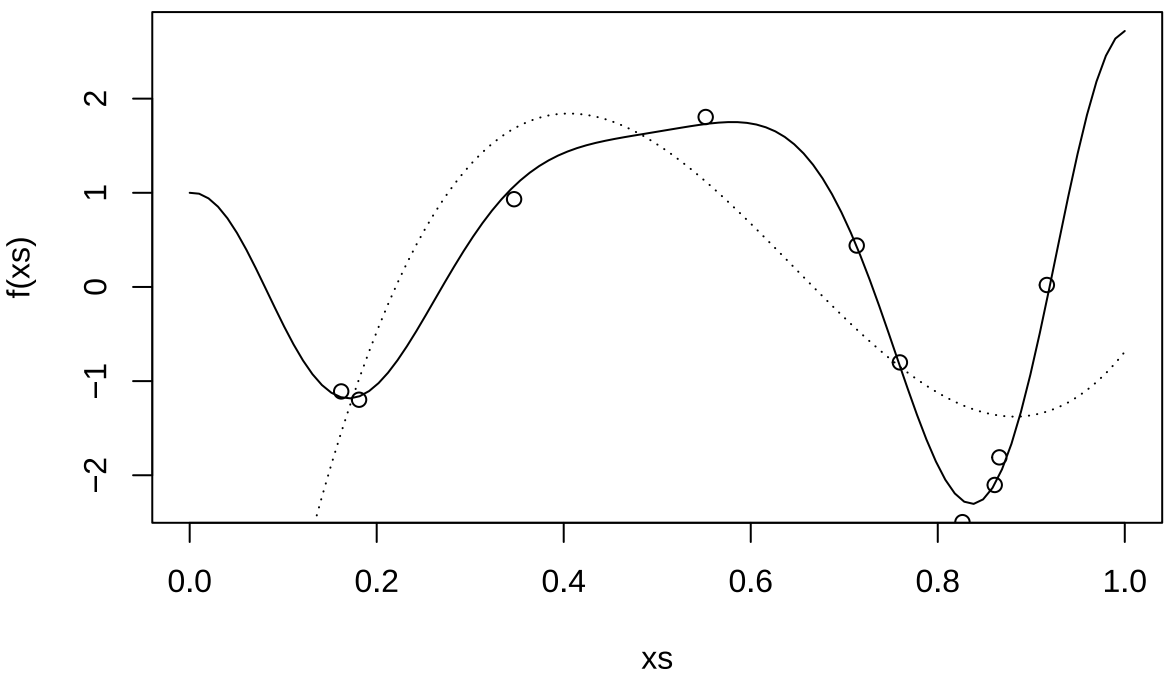
$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$$

3 Méthode des moindres carrés appliquée à la régression polynomiale

Pour un polynôme de degré $N-1$, les fonctions de bases mentionnées ci-dessus sont : $f_1(x) = 1$, $f_2(x) = x$, $f_3(x) = x^2$,..., $f_N(x) = x^{N-1}$. Elles permettent de définir la matrice des données $\mathbf{A}$ et la matrice de Gram $\mathbf{A}^T\mathbf{A}$.

Nous reprenons l'exemple synthétique du précédent chapitre et nous résolvons le système linéaire correspondant à la matrice de Gram pour un polynôme de degré fixé.

```
set.seed(1123)
# Image par f d'un échantillon uniforme sur l'intervalle [0,1], avec ajout d'un
# bruit gaussien de moyenne nulle et d'écart type 0.2
data = gendat(10,0.2)
coef = polyreg2(data,3)
plt(data,f)
pltpoly(coef)
```



Ce polynôme de degré trois modélise mieux la fonction génératrice inconnue que celui de degré quatre qui ne commettait aucune erreur sur les données observées.