

Thème 2

Moindres carrés : OLS-ODR-TLS



Département Sciences du Numérique
Janvier 2019

2.1 La notation «boîte noire»

On considère un système¹ qui se prête à des mesures sur deux types de grandeurs à valeurs dans \mathbb{R} . D'une part, les **covariables** ou **régresseurs** ou variables **explicatives** ou **indépendantes** ou **en entrée** peuvent être fixées et le système réagit à leur valeur. D'autre part, les **réponses** ou variables **expliquées** ou **dépendantes** ou **en sortie** sont telles que le système produit leur valeur, on peut les mesurer. Au système est associé un modèle mathématique paramétré (la fonction f). On souhaite déterminer la valeur des **paramètres** à partir de mesures des variables –**les observations**– réalisées au cours de plusieurs (n) expériences.

Formellement, nous notons :

- les variables en entrée $x \in \mathbb{R}^m$;
- les variables en sortie $y \in \mathbb{R}^q$;
- les paramètres $\beta \in \mathbb{R}^p$;
- le modèle

$$\begin{aligned} f : \mathbb{R}^m \times \mathbb{R}^p &\longrightarrow \mathbb{R}^q \\ (x, \beta) &\longmapsto y \end{aligned}$$

En analyse de données, on parle aussi de modèle de **régression** au sens où la fonction f et ses paramètres permettent une réduction des données d'un phénomène complexe en vue de les représenter par une loi simplificatrice. En régression, nous rassemblons toutes les informations utiles à la modélisation d'un système sous la forme synthétique graphique de la **boîte noire** :

1. Le système peut être physique, biologique, financier, complexe voire même chimique.

$$x \in \mathbb{R}^m \rightarrow \boxed{\begin{array}{c} \beta \in \mathbb{R}^p \\ m, n, p, q \end{array}} \rightarrow y = f(x, \beta) \in \mathbb{R}^q$$

2.2 Modèles de régression et d'erreurs

À cette boîte noire, nous adjoignons le résultat des mesures issues de chaque expérience. Pour l'expérience i ($i \in \llbracket 1, n \rrbracket$) ce sont la valeur \tilde{x}_i de la variable x et la valeur \tilde{y}_i de la variable y . Ce procédé de calcul des β , faisant intervenir n observations des variables x et y est connu sous le nom d'**estimation de paramètres**.

Dire que le modèle est linéaire est un abus de langage pour désigner le cas où f est affine en β . C'est un cas simple et de référence auquel on se ramène souvent. Le modèle peut s'écrire sous la forme

$$f(x, \beta) = C(x)\beta + D(x)$$

avec

$$C : \mathbb{R}^m \longrightarrow \mathbb{R}^{q \times p}$$

$$D : \mathbb{R}^m \longrightarrow \mathbb{R}^q$$

Dans la suite, on supposera pour simplifier le message que $m = q = 1$ (cas unidimensionnel). La mesure des variables est sujette à des erreurs dont le processus d'estimation doit tenir compte. Ces erreurs peuvent apparaître sur les entrées et/ou les sorties.

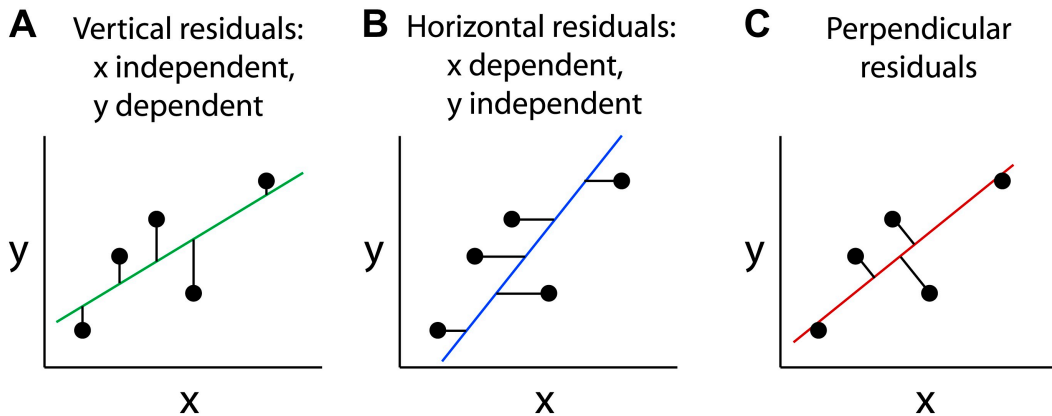


FIGURE 2.1 – Cas de bruitage de données

Les erreurs (additives) sont souvent supposées gaussiennes. Dans tous les cas, deux hypothèses d'indépendance (au sens des variables aléatoires qui les modélisent) sous-tendent ces modèles d'erreur : indépendance d'une expérience à l'autre et indépendance entre toute erreur en entrée et toute erreur en sortie. Les erreurs gaussiennes apparaissent souvent quand des facteurs multiples incontrôlés s'accumulent pour perturber faiblement la mesure.

Dans la suite, on considère des erreurs sur les sorties uniquement. On perturbe la valeur exacte y_i selon $\tilde{y}_i = y_i + \varepsilon_i$.

2.3 Moindres carrés ordinaires

Ces méthodes cherchent à rendre « aussi vraie que possible » l'approximation sous-jacente au modèle de régression. Pour cela, elles calculent β qui minimise « conjointement » tous les

$$\|\tilde{y}_i - f(x_i, \beta)\|^2 \quad \forall i \in \llbracket 1, n \rrbracket$$

Cette expression basique des **résidus** sera plus tard modifiée pour intégrer une pondération des mesures et prendre en compte les erreurs sur les entrées.

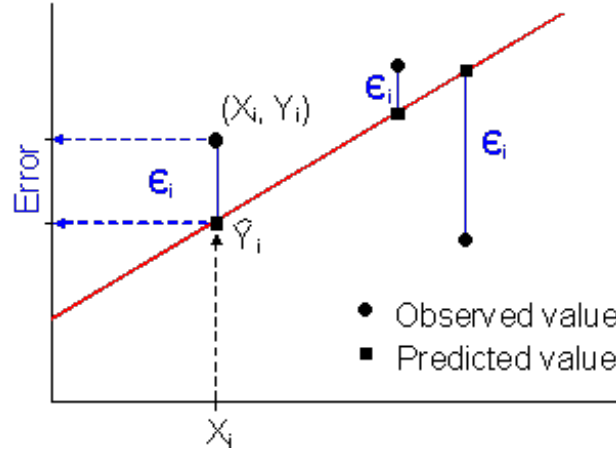


FIGURE 2.2 – Moindres carrés ordinaires

Moindres carrés ordinaires (*Ordinary Least Squares*, OLS)

On parle aussi de « régression en distance verticale ». L'estimateur est :

$$\widehat{\beta}_{\text{OLS}} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^n \|\tilde{y}_i - f(x_i, \beta)\|^2 \quad (2.1)$$

Exemple : On souhaite modéliser le problème consistant à trouver la transformation affine liant deux ensembles de points dans \mathbb{R}^2 .

$$\begin{bmatrix} u'_i \\ v'_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} + \begin{bmatrix} a_3 \\ b_3 \end{bmatrix}$$

Modélisation par "boîte noire", on cherche un vecteur $\beta \in \mathbb{R}^6$ tels que on puisse exprimer les coordonnées des points $X' = (u', v')$ en fonction des points $X = (u, v)$. En posant $\beta = [a_1 \ a_2 \ a_3 \ b_1 \ b_2 \ b_3]^T$, identifier A et b tel qu'on puisse formuler ce problème en moindres carrés ordinaires :

$$\min_{\beta \in \mathbb{R}^6} \frac{1}{2} \|A\beta - b\|_2^2$$

2.3.1 Résolution par pseudo inverse

Si f est linéaire en β , le problème OLS se ramène à :

$$\widehat{\beta_{\text{OLS}}} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|A\beta - b\|^2 \quad (2.2)$$

où

$$A = \begin{bmatrix} C(\tilde{x}_1) \\ \vdots \\ C(\tilde{x}_n) \end{bmatrix} \in \mathbb{R}^{n \times p} \text{ et } b = \begin{bmatrix} \tilde{y}_1 - D(\tilde{x}_1) \\ \vdots \\ \tilde{y}_n - D(\tilde{x}_n) \end{bmatrix} \in \mathbb{R}^n$$

Une méthode de résolution des équations normales, vue en cours, utilise la pseudo inverse A^+ de A (que l'on peut exprimer via la SVD).

$$\widehat{\beta_{\text{OLS}}} = A^+ b$$

2.3.2 Résolution par moindres carrés totaux (TLS, Total Least Squares)

Formulation implicite

Avec la formulation implicite, les sorties disparaissent de la boîte noire. Les relations entre variables sont du type $f(x, \beta) = 0$.

Exemple : Soit un ensemble de n points de \mathbb{R}^2 , on cherche à trouver les coefficients de la droite qui approche le mieux ce nuage de points : $ax + by + c = 0$.

On note $\beta \in \mathbb{R}^3$ le vecteur de paramètres défini par : $\beta = [a \ b \ c]^T$. En normalisant, on obtient :

$$\frac{a}{\sqrt{a^2 + b^2 + c^2}}x + \frac{b}{\sqrt{a^2 + b^2 + c^2}}y + \frac{c}{\sqrt{a^2 + b^2 + c^2}} = 0$$

Ecriture matricielle :

$$\left\{ \begin{array}{l} \begin{bmatrix} x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \\ \|\beta\| = 1 \end{array} \right.$$

On se ramène au problème suivant :

$$\min_{\|\beta\|=1} \frac{1}{2} \|A\beta\|_2^2$$

L'idéal serait de trouver $\beta \neq 0_{\mathbb{R}^3}$ tel que $A\beta = 0$ c'est-à-dire trouver β un élément du noyau de A .

Résolution de moindres carrés ordinaires :

Une formulation aux moindres carrés totaux est également possible grâce à la SVD. Il s'agit de résoudre :

$$\min_{\beta \in \mathbb{R}^p} \|A\beta - b\|^2$$

Rappel sur la décomposition en valeurs singulières SVD

Soit $A \in \mathbb{R}^{q \times p}$ ($q \geq p$) alors on peut décomposer A :

$$A = U\Sigma V^T$$

avec :

- $U \in \mathbb{R}^{q \times q}$ est formée de q vecteurs propres orthonormés associés aux q valeurs propres de AA^T .
- $V \in \mathbb{R}^{p \times p}$ est formée de p vecteurs propres orthonormés associés aux p valeurs propres de $A^T A$.
- $\Sigma \in \mathbb{R}^{q \times p}$ est une matrice rectangulaire dont les éléments non nuls sur la diagonale sont les valeurs singulières $\sigma_i, i = \{1, \dots, q\}$ de A (sont les racines carrées des valeurs propres de $A^T A$ et de AA^T).

Propriétés :

- $\text{rang}(A) = p$ donc $\sigma_{p+1} = \dots = \sigma_q = 0$.
- Si $A = U\Sigma V^T$ alors $A^T = V\Sigma U^T$ est une SVD de A^T .

On peut réécrire le problème $\frac{1}{2}\|A\beta - b\|_2^2$ en un système augmenté :

$$A\beta - b = [A \ b] \begin{bmatrix} \beta \\ -1 \end{bmatrix} = 0$$

On cherche donc un élément du noyau de $[A \ b]$ de la forme $\begin{bmatrix} \beta \\ -1 \end{bmatrix}$. Si le rang de $[A \ b]$ est $p+1$, on passe au rang p :

$$[\hat{A} \ \hat{b}] = [A \ b] - \sigma_{p+1} u_{p+1} v_{p+1}^T$$

où u_{p+1}, v_{p+1} et σ_{p+1} proviennent de la SVD de $[A \ b]$.

$\widehat{\beta}_{TLS}$ est associé à v_{p+1} mais doit être de la forme $\begin{bmatrix} \beta \\ -1 \end{bmatrix}$.

Soit $v_{p+1} = \begin{bmatrix} v_{1 \ p+1} \\ \vdots \\ v_{p+1 \ p+1} \end{bmatrix} \in \mathbb{R}^{p+1}$ d'où

$$\widehat{\beta}_{TLS} = \frac{1}{-v_{p+1 \ p+1}} \begin{bmatrix} v_{1 \ p+1} \\ \vdots \\ v_{p \ p+1} \end{bmatrix}.$$

2.4 Exercice : transformations géométriques 2D

Les opérations géométriques ont pour but de modifier la position des informations contenues dans l'image sans modifier le niveau de gris. Ces opérations peuvent s'appliquer à la totalité des points d'une image (les pixels), un objet particulier de l'image, voire à certains points spécifiques (recalage de points caractéristiques).

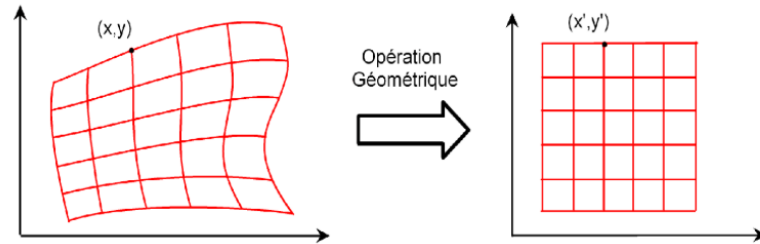


Fig 1 - Recalage géométrique d'une image

Les opérations de base permettent de réaliser des transformations géométriques simples. Elles sont visibles lorsqu'elles affectent une structure spécifique (élément carré, maillage). On considère la transformation suivante :

Cisaillement double /transvection		$\begin{cases} x' = x + \alpha y \\ y' = \beta x + y \end{cases}$
--	--	---

Questions

On souhaite modéliser le problème consistant à trouver la transvection liant 2 ensembles de points dans \mathbb{R}^2 . On définit $\gamma \in \mathbb{R}^2$ le vecteur de paramètres inconnus défini par : $\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

$$U = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \longrightarrow \boxed{\gamma \in \mathbb{R}^2} \longrightarrow U' = \begin{bmatrix} x' \\ y' \end{bmatrix} = f(U, \gamma) \in \mathbb{R}^2$$

1. Déterminer la matrice $C \in \mathbb{R}^{2 \times 2}$ et le vecteur $d \in \mathbb{R}^2$ de la fonction $f(U, \gamma)$ définie sous la forme :

$$f(U, \gamma) = C \gamma + d$$

2. Ecrire le problème d'optimisation pour définir le vecteur de paramètres γ en considérant un ensemble de n points transformés par la même transvection.
3. Calculer la solution théorique de ce problème d'optimisation.

2.5 ODR et Moindres carrés totaux

On considère désormais des erreurs à la fois sur les données d'entrée et sur des données sorties. On note $i \in \llbracket 1, n \rrbracket$ l'indice de l'expérience. Les erreurs sur les entrées perturbent la valeur exacte x_i $\tilde{x}_i = x_i + \delta_i$ sont complétées par des erreurs sur les sorties : $\tilde{y}_i = y_i + \varepsilon_i$.

Moindres carrés orthogonaux (*Orthogonal Distance Regression*, ODR)

Les entrées et les sorties sont entachées d'erreurs. Il s'agit d'estimer « conjointement » β et les erreurs en entrée δ_i :

$$P_{\text{ODR}} \left\{ \begin{array}{l} \min \sum_{i=1}^n \|\tilde{y}_i - f(\tilde{x}_i - \delta_i, \beta)\|^2 + \|\delta_i\|^2 \\ \beta \in \mathbb{R}^p, \delta_i \in \mathbb{R}^m, i \in \llbracket 1, n \rrbracket \end{array} \right. \quad (2.3)$$

On parle de « régression en distance orthogonale » (*Orthogonal Distance Regression*, ODR).

Moindres carrés totaux

Le problème ODR admet une solution analytique dans le cadre des moindres carrés totaux (*Total Least Squares*), TLS. C'est un cas restreint où les dimensions vérifient $m = p$ et $q = 1$ et le modèle f est bilinéaire de la forme $f(x, \beta) = x^\top \beta + q_0^\top \beta + q_1$ avec $q_0 \in \mathbb{R}^p$ et $q_1 \in \mathbb{R}$. Si $f(\tilde{x}_i - \delta_i, \tilde{\beta}) = \tilde{y}_i - \varepsilon_i, \forall i \in \llbracket 1, n \rrbracket$ alors en exploitant les linéarités et en synthétisant les égalités sous forme matricielle :

$$\left(\underbrace{\begin{bmatrix} \tilde{x}_1^\top + q_0^\top \\ \vdots \\ \tilde{x}_n^\top + q_0^\top \end{bmatrix}}_A + \underbrace{\begin{bmatrix} -\delta_1^\top \\ \vdots \\ -\delta_n^\top \end{bmatrix}}_E \right) \tilde{\beta} = \underbrace{\begin{bmatrix} \tilde{y}_1 - q_1 \\ \vdots \\ \tilde{y}_n - q_1 \end{bmatrix}}_b + \underbrace{\begin{bmatrix} -\varepsilon_1 \\ \vdots \\ -\varepsilon_n \end{bmatrix}}_r$$

Par suite, l'équation (2.3) s'écrit :

$$P_{\text{TLS}} \left\{ \begin{array}{l} \min \|[E \ r]\|^2 \\ \beta \in \mathbb{R}^p, E \in \mathbb{R}^{n \times p}, r \in \mathbb{R}^n \\ (A + E)\beta = b + r \end{array} \right.$$

La résolution de (P_{TLS}) se base sur la décomposition en valeurs singulières (SVD) de $[A \ b]$.