Moindres carrés

Vincent Nozick

Les moindres carrés

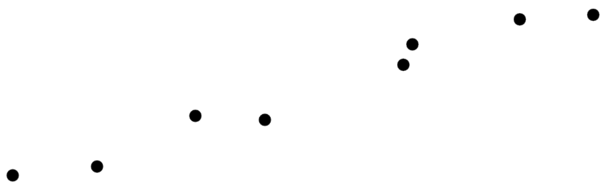
Introduction :

On dispose d'un ensemble de point supposés alignés. On cherche l'équation de la droite qui représente le mieux cet ensemble de points.

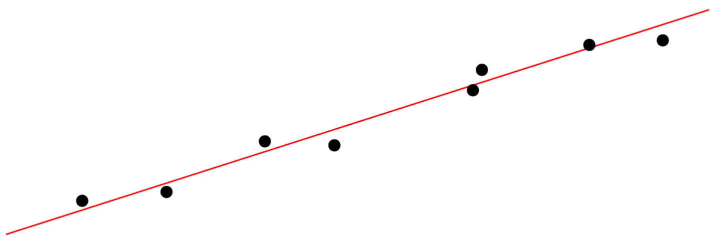
Problème :

Les points ne sont en général pas tout à fait alignés.

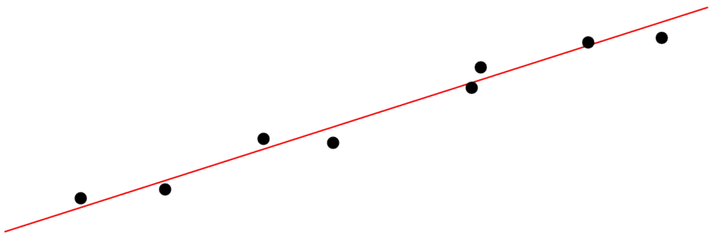
Approximation d'une droite



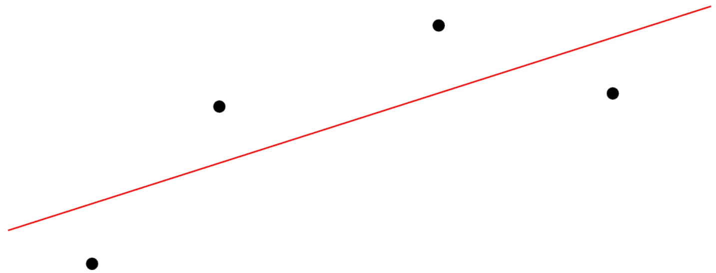
Approximation d'une droite



Approximation d'une droite

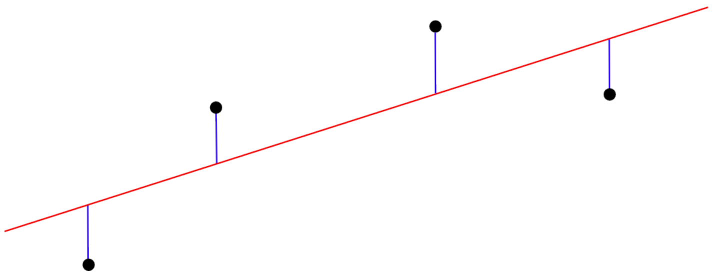


Approximation d'une droite



$y_{th} = a.x + b$

Approximation d'une droite



$y_{th} = a.x + b$

$$R^2 = \sum_{i=1}^{N} \Big(y_{th}(x_i) - y_i \Big)^2 = \sum_{i=1}^{N} (ax_i + b - y_i)^2$$

Le résidu

$$R^2 = \sum (ax_i + b - y_i)^2$$

Le résidu

$$R^2 = \sum (ax_i + b - y_i)^2$$

$$R^2 = \sum \left(a^2x_i^2 + b^2 + y_i^2 + 2abx_i - 2ax_iy_i - 2by_i\right)$$

Approximation d'une droite | Moindres carrés et matrices | Exemple | Applications

Le résidu

$$R^2 = \sum (ax_i + b - y_i)^2$$

$$R^2 = \sum \left(a^2x_i^2 + b^2 + y_i^2 + 2abx_i - 2ax_iy_i - 2by_i\right)$$

$$R^2 = a^2 \sum x_i^2 + b^2 \sum 1 + \sum y_i^2 + 2ab \sum x_i - 2a \sum x_iy_i - 2b \sum y_i$$

Vincent Nozick | Moindres carrés | 5 / 24

Approximation d'une droite | Moindres carrés et matrices | Exemple | Applications

Dérivée partielle du résidu

$$R^2 = a^2 \sum x_i^2 + b^2 \sum 1 + \sum y_i^2 + 2ab \sum x_i - 2a \sum x_iy_i - 2b \sum y_i$$

On veut minimiser $R^2 \to$ on annule ses dérivées

$$\frac{\partial R^2}{\partial a} = 0 \quad \Rightarrow \quad 2a \sum x_i^2 + 2b \sum x_i - 2 \sum x_iy_i = 0$$

$$a \sum x_i^2 + b \sum x_i = \sum x_iy_i$$

Approximation d'une droite | Moindres carrés et matrices | Exemple | Applications

Dérivée partielle du résidu

$$R^2 = a^2 \sum x_i^2 + b^2 \sum 1 + \sum y_i^2 + 2ab \sum x_i - 2a \sum x_iy_i - 2b \sum y_i$$

On veut minimiser $R^2 \to$ on annule ses dérivées

$$\frac{\partial R^2}{\partial b} = 0 \quad \Rightarrow \quad 2bN + 2a \sum x_i - 2 \sum y_i = 0$$

$$a \sum x_i + bN = \sum y_i$$

Système linéaire

$$a\sum x_i^2 + b\sum x_i = \sum x_i y_i$$

$$a\sum x_i + bN = \sum y_i$$

On obtient un système linéaire :

$$\left\{\begin{array}{lll} a\sum x_i^2 & +b\sum x_i & = \sum x_i y_i \\ a\sum x_i & +bN & = \sum y_i \end{array}\right.$$

Système linéaire

Solution :

$$a = \frac{\sum_{i=1}^N x_i \sum_{i=1}^N y_i - N\sum_{i=1}^N x_i y_i}{\left(\sum_{i=1}^N x_i\right)^2 - N\sum_{i=1}^N x_i^2}$$

$$b = \frac{\sum_{i=1}^N x_i y_i \sum_{i=1}^N x_i - \sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i}{\left(\sum_{i=1}^N x_i\right)^2 - N\sum_{i=1}^N x_i^2}$$

Moindres carrés et matrices

Le système linéaire de l'approximation d'une droite :

$$\left\{\begin{array}{lll} a\sum x_i^2 & +b\sum x_i & = \sum x_i y_i \\ a\sum x_i & +bN & = \sum y_i \end{array}\right.$$

peut s'écrire matriciellement

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & N \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

Moindres carrés et matrices

Une autre approche plus intuitive serait de résoudre :

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

On cherche a et b tels que $y_i = ax_i + b \quad \forall i \in 1...n$

Système linéaire

Pour un système linéaire $\mathtt{A}\mathbf{x} = \mathbf{b}$ ($\mathtt{A}$: m lignes, n colonnes)

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Système linéaire

Pour un système linéaire $\mathtt{A}\mathbf{x} = \mathbf{b}$ ($\mathtt{A}$: m lignes, n colonnes)

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

- si $m = n$ (et $\mathtt{A}$ de rang plein) : **solution unique**.
- si $m < n$: **infinité de solutions**.
- si $m > n$ (et $\text{rang}(\mathtt{A}) = n$) : **généralement pas de solution**.

Système surdéterminé

si $m > n$ (et $rang(\mathtt{A}) = n$)
$\rightarrow$ on peut trouver une solution au sens des moindres carrés.

Autrement dit :
On a plus d'équations que d'inconnues, qui ne sont pas d'accord entre elles. On peut chercher une solution satisfaisant au mieux toutes les équations du système.

équation normale et pseudo-inverse

Dans le système surdéterminé $\mathtt{A}_{m\times n}\mathbf{x} = \mathbf{b}$, $m > n$

$\rightarrow$ aucune solution $\mathbf{x}$ ne peut satisfaire le système.
$\rightarrow$ on cherche une solution $\mathbf{x}$ telle que :

$$e(\mathbf{x}) = \|\mathtt{A}\mathbf{x} - \mathbf{b}\|_2^2 \qquad \text{soit minimal.}$$

équation normale et pseudo-inverse

On veut minimiser : $e(\mathbf{x}) = \|\mathtt{A}\mathbf{x} - \mathbf{b}\|^2$

$\rightarrow$ on cherche $\mathbf{x}$ tel que $\dfrac{\partial e(\mathbf{x})}{\partial \mathbf{x}} = 0$

$$\frac{\partial \|\mathtt{A}\mathbf{x} - \mathbf{b}\|^2}{\partial \mathbf{x}} = 2\mathtt{A}^\top(\mathtt{A}\mathbf{x} - \mathbf{b}) = 0 \quad \rightarrow \quad \mathtt{A}^\top\mathtt{A}\mathbf{x} - \mathtt{A}^\top\mathbf{b} = 0$$

Équation normale

$$\mathtt{A}^\top\mathtt{A}\mathbf{x} = \mathtt{A}^\top\mathbf{b}$$

solution (au sens des moindres carrés)

$$\Rightarrow \mathbf{x} = (\mathtt{A}^\top\mathtt{A})^{-1}\mathtt{A}^\top\mathbf{b}$$

Matrice pseudo-inverse

$$\mathbf{x} = \underbrace{(\mathtt{A}^\top\mathtt{A})^{-1}\mathtt{A}^\top}_{\mathtt{A}^+}\mathbf{b}$$

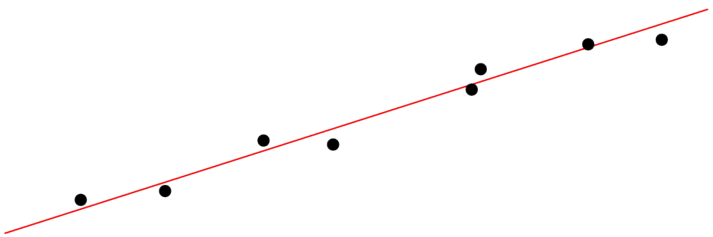
Matrice pseudo inverse : $\mathtt{A}^+ \overset{\text{def}}{=} (\mathtt{A}^\top\mathtt{A})^{-1}\mathtt{A}^\top$

Propriétés :

- si $rang(\mathtt{A}) = n$, alors $\mathtt{A}^\top\mathtt{A}$ est une matrice carrée régulière (et $\mathtt{A}^+$ est donc tout le temps défini).
- $\mathtt{A}^+ = \mathtt{A}^{-1}$ si $\mathtt{A}$ est régulière (carrée de rang plein).

Matrice pseudo-inverse

Exemple : Approximation d'une droite



Trouver la droite $y = ax + b$
passant au mieux par les points (x_i, y_i).

Matrice pseudo-inverse

On cherche a et b tels que $y_i = ax_i + b \quad \forall i \in 1...n$

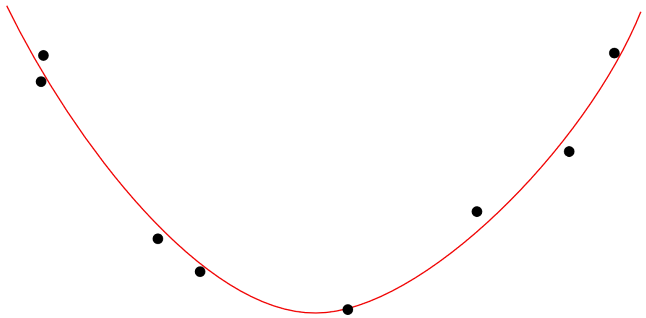
$$\rightarrow \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}}_{\mathtt{U}} \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{v}}$$

Il s'agit bien d'un système surdéterminé, dont la solution est :

$$\begin{pmatrix} a \\ b \end{pmatrix} = \mathtt{U}^+\mathbf{v}$$

Matrice pseudo-inverse

Exemple : Approximation d'une parabole



Trouver le polynôme $y = ax^2 + bx + c$
passant au mieux par les points (x_i, y_i).

Matrice pseudo-inverse

On cherche a, b et c tels que $y_i = ax_i^2 + bx_i + c \quad \forall i \in 1...n$

$$\rightarrow \underbrace{\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ \vdots & \vdots & \\ x_n^2 & x_n & 1 \end{bmatrix}}_{\mathtt{U}} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{v}}$$

Il s'agit bien d'un système surdéterminé, dont la solution est :

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathtt{U}^{+}\mathbf{v}$$

Les moindres carrés

A retenir :

Globalement, les moindres carrés permettent de résoudre des systèmes surdéterminés (où il y a plus d'équations que d'inconnues).

Les moindres carrés

Applications :

- Physique
- Traitement d'image ou du signal
- Biologie
- Économie
- ...

La régression linéaire

Variable à expliquer :
le salaire y_i d'un panel de personnes interrogées.
(en pratique on prend le $\log$(salaire))

Variable explicative
on cherche à voir si les variable suivante expliquent ce salaire :

- $\mathbf{x_1}$: nombre d'années d'études
- $\mathbf{x_2}$: nombre d'années d'expérience professionnelle
- $\mathbf{x_3}$: la personne est une femme (1=oui, 0=non)

La régression linéaire

Formulation :

$$\begin{pmatrix} \vdots \\ y_i \\ \vdots \end{pmatrix} = \underbrace{\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^i & x_2^i & x_3^i \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}}_{\mathtt{A}} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \vdots \\ u_i \\ \vdots \end{pmatrix}$$

où $\mathbf{u}$ représente le résidu généré par les variables non observées.

Solution :

$$\beta = \mathtt{A}^+ \mathbf{y}$$

$$\mathbf{u} = \mathbf{y} - \mathtt{A}\beta$$