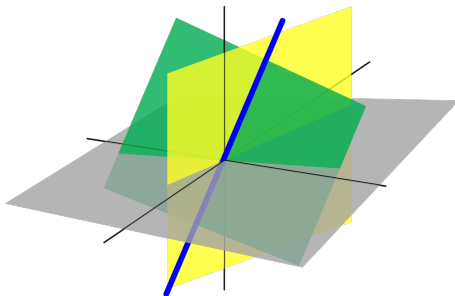


ALGÈBRE LINÉAIRE

MICROTECHNIQUE

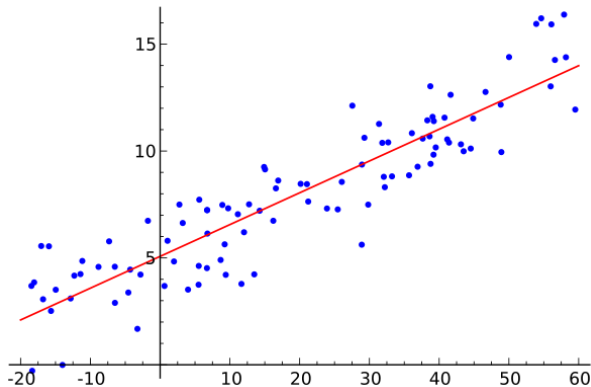
RÉGRESSION LINÉAIRE

Jérôme Scherer



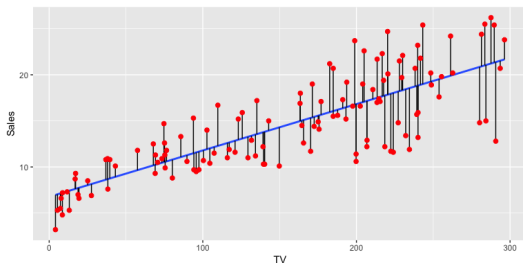
OBJECTIF

On se donne un “nuage” de points dans le plan, donnés par leurs coordonnées $(x_1, y_1), \dots, (x_n, y_n)$ et on aimerait trouver la droite qui donne la meilleure approximation :



FORMALISATION

On cherche la droite d'équation $y = ax + b$ la plus proche des points $(x_1, y_1), \dots, (x_n, y_n)$ dans le sens où les distances verticales entre les points et la droite sont minimisées (voici un exemple tiré de UC Business Analytics R Programming Guide) :



Autrement dit les distances $|ax_n + b - y_n|$ doivent être les plus petites possibles.

LES MOINDRES CARRÉS

Pour ne pas devoir discuter de valeurs absolues, on minimise les carrés de ces distances. On cherche donc les nombres a et b tels que

$$\sum_{i=1}^n (ax_n + b - y_n)^2$$

est minimal. Par conséquent le système suivant de n équations

$$\begin{cases} ax_1 + b = y_1 \\ \dots \dots \\ ax_n + b = y_n \end{cases}$$

est incompatible (à moins d'un coup de chance) et on cherche la meilleure solution $(\hat{a}, \hat{b})$ au sens des moindres carrés !

FORME MATRICIELLE

on cherche a, b

On écrit le système ci-dessus sous forme matricielle

$$\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

et on utilise l'équation normale pour résoudre :

$$\begin{pmatrix} \sum (x_i)^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

REMARQUE

Toutes les sommes sont indicées par un entier i entre 1 et n .

LA SOLUTION

REMARQUE

En général, dans la pratique, les colonnes de x_i et celle de 1 ne sont pas proportionnelles dans la matrice A , la solution est alors unique : Le déterminant de la matrice $A^T A$ est non nul.

Ainsi $D = n \sum (x_i)^2 - (\sum x_i)^2 \neq 0$. Alors

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \frac{1}{D} \begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum (x_i)^2 \end{pmatrix} \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

THÉORÈME

La solution au sens des moindres carrés est donnée par

$$\hat{a} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{D} \text{ et}$$
$$\hat{b} = \frac{[\sum (x_i)^2](\sum y_i) - (\sum x_i)(\sum x_i y_i)}{D}.$$