

Exercice – Droite de régression linéaire

Etant donné un entier n supérieur ou égal à 2, on considère un nuage de n points du plan, de coordonnées $(x_1, y_1), \dots, (x_n, y_n)$. On suppose que les réels $x_1, x_2, \dots, x_n$ (resp. $y_1, y_2, \dots, y_n$) ne sont pas tous égaux.

L'objectif de l'exercice est de trouver la droite qui approche "au mieux" ces différents points.

Soit f la fonction définie sur $\mathbb{R}^2$ à valeurs réelles qui, à tout couple (a, b) de $\mathbb{R}^2$, associe le réel

$$f(a, b) = \sum_{i=1}^n (ax_i + b - y_i)^2.$$

f est la somme des carrés des distances des points à la droite. On cherche à trouver un minimum de cette fonction.

Si $x = (x_1, \dots, x_n)$, on rappelle que la moyenne arithmétique $\overline{x}$ et l'écart-type σ_x de sont définis par : $\overline{x} =$

$$\frac{1}{n} \sum_{i=1}^n x_i \quad \sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^2}$$

On rappelle enfin que la covariance $\text{cov}(x, y)$ du couple (x, y) est définie par :

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$$

1. Notons $S = \sum_{i=1}^n x_i$, $T = \sum_{i=1}^n x_i^2$ et $U = \sum_{i=1}^n x_i y_i$

a) Exprimer S en fonction de n et $\overline{x}$.

b) Exprimer T en fonction de n , $\overline{x}$ et σ_x .

c) Exprimer U en fonction de n , $\overline{x}$, $\overline{y}$ et $\text{cov}(x, y)$.

2. Justifier que f est de classe C^2 sur $\mathbb{R}^2$.

3. a) Ecrire le système d'équations (S) permettant de déterminer les points critiques de f .

b) Résoudre le système. En déduire que f admet un unique point critique (a_0, b_0) que l'on exprimera en fonction de $\overline{x}$, $\overline{y}$, σ_x^2 et $\text{cov}(x, y)$.

c) Montrer que ce point critique est un minimum local de la fonction.