LA MODÉLISATION DES DONNÉES PAR LA RÉGRESSION FREINE

RÉSUMÉ DE LA THÉORIE ET DES APPLICATIONS

Régression linéaire

Un ensemble donné de paires de valeurs est considéré :(x_1,y_1), (x_2,y_2), ... (x_n,y_n), L'exemple le plus simple d'une fonction continue qui se rapproche de la dépendance y(x), basée sur des paires discrètes connues, est une ligne droite $f(x) = a_0 + a_1 x$. Méthode d'approximation dans le sens des plus petits carrés, le droit de régression est uniquement déterminé par la condition de minimiser la somme des erreurs soulevées à la deuxième puissance.

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 .$$

L'annulation des dérivés partiels de la Sum S aux coefficients et conduit à un système d'équations avec la solution: a_0 a_1

$$\begin{cases} a_1 = \dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\ a_0 = \bar{y} - a_1 \bar{x} \end{cases}$$

où sont les moyennes arithmétiques des valeurs y et $\bar{y}$ si $\bar{x}_i$ respectivement, x_i.

L'écart standard entre les valeurs connues et les bonnes valeurs est

$$\sigma_y = \sqrt{\frac{S_r}{n-2}} .$$

Régression linéaire multiple

Dans le cas de la dépendance à une taille y de deux paramètres x_1 et x_2, la fonction d'approximation la plus simple est l'expression linéaire $f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2$.

L'optimisation de la somme des carrés d'erreur permet la détermination sans ambiguïté des coefficients a_0, a_1 si a_2:

$$\begin{cases} n a_0 + a_1 \sum_{i=1}^{n} x_{1,i} + a_2 \sum_{i=1}^{n} x_{2,i} = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_{1,i} + a_1 \sum_{i=1}^{n} x_{1,i}^2 + a_2 \sum_{i=1}^{n} x_{1,i} x_{2,i} = \sum_{i=1}^{n} x_{1,i} y_i \\ a_0 \sum_{i=1}^{n} x_{2,i} + a_1 \sum_{i=1}^{n} x_{1,i} x_{2,i} + a_2 \sum_{i=1}^{n} x_{2,i}^2 = \sum_{i=1}^{n} x_{2,i} y_i \end{cases}$$

Régressions liniarizată

Si le modèle mathématique de la foule point est une fonction non linéaire, parfois il peut se transformer en l'équation d'une droite à travers les opérations mathématiques.

Par exemple, une dépendance exponentielle du type $y = a_1 e^{b_1 x}$, oùun 1 et b_1 sont constants, sont transformés par logarithm $\ln y = \ln a_1 + b_1 x$. Si noté $y^* = \ln y$, $a_0^* = \ln a_1$ şi $a_1^* = b_1$, ensemble de données approximatives avec (x^*, y^*) $f(x^*) = a_0^* + a_1^* \cdot x^*$. Une fois que vous apprenez ces coefficients à partir de l'application de l'algorithme de régression linéaire, les coefficients initiaux:

$a_1 = e^{a_0^*}$ şi $b_1 = a_1^*$.

Aplicaţia 1.

Trouver l'équation d'une ligne droite qui se rapproche, dans le sens des plus petits carrés, l'ensemble de données indiqué dans le tableau suivant et calculer l'écart standard (déviation) de l'approximation:

i	1	2	3	4	5	6	7
x_i	1	2	3	4	5	6	7
y_i	0.5	2.5	2	4	3.5	6	5.5

Aplicaţia 2.

Utiliser une régression linéaire multiple pour approximer les données dans le tableau suivant:

x_1	0	2	2,5	1	4	7
x_2	0	1	2	3	6	2
y	5,1	10,2	8,9	0	2,8	27,2

Indicaţie:

Résoudre le système, les coefficients sont obtenusa a_0, a_1 şi a_2.

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{1,i} & \sum_{i=1}^{n} x_{2,i} \\ \sum_{i=1}^{n} x_{1,i} & \sum_{i=1}^{n} x_{1,i}^2 & \sum_{i=1}^{n} x_{1,i}x_{2,i} \\ \sum_{i=1}^{n} x_{2,i} & \sum_{i=1}^{n} x_{1,i}x_{2,i} & \sum_{i=1}^{n} x_{2,i}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1,i} y_i \\ \sum_{i=1}^{n} x_{2,i} y_i \end{Bmatrix}$$

Aplicaţia 3.

Le flux volumétrique à travers un tuyau dépend de son diamètre et de sa pente selon une loi du type:. $V = a_0 \cdot D^{a_1} \cdot P^{a_2}$ Trouvez le débit correspondant à un tuyau avec D-25 cm et P-0.025 si les données expérimentales suivantes sont connues:

D=[10 20 30 10 20 30 10 20 30]; [cm]

P=[0.001 0.001 0.001 0.01 0.01 0.01 0.05 0.05 0.05]; [-]

V=[140 830 2420 470 2890 8400 1110 6900 20000]; [cm^3/sec]

Indicaţie:

It is liniarizeaz-rela-aia de depends:

$\log(V) = \log(a_0) + a_1 \cdot \log(D) + a_2 \cdot \log(P)$

$x_1 = \log(D)$;

$x_2 = \log(P)$;

y=log(V);

De minimiser la somme des carrés d'erreur, trois équations linéaires dans les inconnus b_0=log(a_0), a_1 a and $_2$.

Probleme propuse

1.En utilisant la méthode des plus petits carrés, approximatif le droit (fonction degré 1) qui se rapproche des données expérimentales suivantes:

X= [1 2 3 4 5]; Y=[0 4 3 6 8].

2.Utilisez la méthode des plus petits carrés pour approximer à travers une parabole $f(x) = a_o + a_1 x_i + a_2 x_i^2$ les données expérimentales suivantes :

x_i	1	3	8	5
y_i	9	6	9	3

3.Déterminer la fonction $f(x) = a \cdot e^{b/x}$ qui se rapproche dans le sens des plus petits carrés, les valeurs dans le tableau suivant. Spécifier l'écart maximal absolu et relatif. Représenter graphiquement les points donnés et la fonction déterminée.

i	x_i	y_i
0	2	20
1	6	38
2	12	49
3	15	51
4	30	61
5	46	70

4.Utiliser une régression linéaire multiple $f(x_1, x_2)$ pour approximer les données suivantes:

x_1	0	1	2	0	1	2
x_2	2	2	4	4	6	6
y	19	12	11	24	22	15