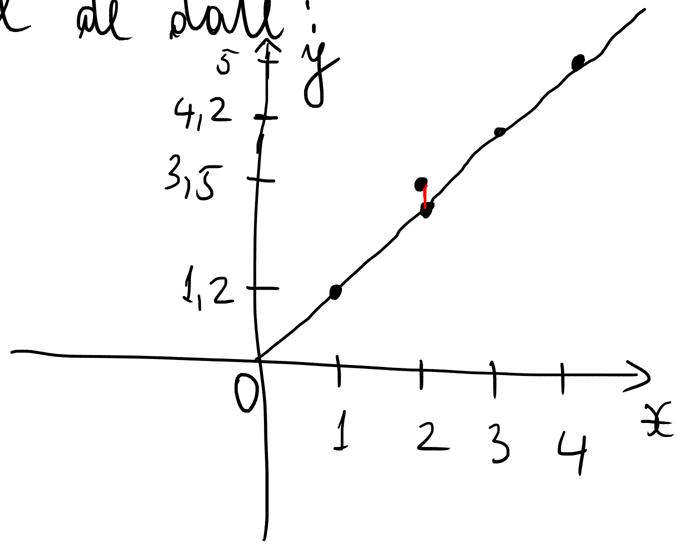
Curs 13

Curbe de regresie

1. Regresia liniară

Considerăm următorul tabel de date:

x_i	y_i
1	1,3
2	3,5
3	4,2
4	5



Conform graficului se observă că relația dintre $x=(x_1, x_2, x_3, x_4)$ și $y=(y_1, y_2, y_3, y_4)$ este liniară. Un motiv pentru care punctele nu se află pe aceeași dreaptă este că datele au fost obținute cu marjă de eroare.

Vom construi o dreaptă de ecuație $y = ax+b$ care trece cel mai aproape de punctele date. Coeficienții a și b sunt necunoscuți și urmează să fie determinați.

Distanța (pe verticală) dintre punctul (x_i, y_i) și punctul situat pe dreaptă (x_i, ax_i+b) este

$$\sqrt{(x_i-x_i)^2+(y_i-(ax_i+b))^2} = \sqrt{(y_i-(ax_i+b))^2} = |y_i-(ax_i+b)|.$$

Fie $y=(y_1,\dots,y_n)$ și $\overline{y}=(ax_1+b,\dots,ax_n+b)$.

Coeficienții a și b se obțin minimizând funcția $E(a,b)=\|y-\overline{y}\|^2=\sum_{i=1}^{n}\left(y_i-(ax_i+b)\right)^2$ (Reamintim că $\underset{\substack{\| \\ (\alpha_1,\dots,\alpha_n)}}{\|\alpha\|}=\left(\sum_{i=1}^{n}\alpha_i^2\right)^{\frac{1}{2}}$).

Minimizând funcția E se obține poziția optimă a dreptei $y=ax+b$ față de punctele (x_i,y_i), cu $i=\overline{1,n}$.

Funcția E admite un punct de minim care este și punct critic (sau staționar). Rezolvăm sistemul:

$$\begin{cases}\frac{\partial E}{\partial a}(a,b)=0\\ \frac{\partial E}{\partial b}(a,b)=0\end{cases} \Leftrightarrow \begin{cases}2\sum_{i=1}^{n}\left(y_i-(ax_i+b)\right)\cdot(-x_i)=0 \,|:2\\ 2\sum_{i=1}^{n}\left(y_i-(ax_i+b)\right)\cdot(-1)=0\,|:2\end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases}a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_iy_i\\ a\sum_{i=1}^{n}x_i+b\underbrace{\sum_{i=1}^{n}1}_{=n}=\sum_{i=1}^{n}y_i.\end{cases}$$

Am obținut un sistem liniar cu necunoscutele a și b.

Matricea acestui sistem este $A = \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \underbrace{\sum_{i=1}^{n} 1}_{= n} \end{pmatrix}$.

Vectorul termenilor liberi este $w = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}$.

Avem $A \cdot \begin{pmatrix} a \\ b \end{pmatrix} = w$,

Pentru rezolvarea numerică a sistemului de mai sus putem folosi una dintre metodele de eliminare Gauss.

2. <u>Regresia polinomială</u>

Dacă punctele (x_i, y_i), cu $i=\overline{1,n}$ descriu o funcție polinomială, atunci curba care aproximează datele o putem alege de forma unui polinom de grad m,

$$P_m(x) = a_0 + a_1 x + \dots + a_m x^m.$$

Considerăm, în continuare, cazul $m=2$, care corespunde regresiei pătratice.

Fie $P_2(x) = a x^2 + b x + c$.

Minimizând funcția $E(a,b,c) = \sum_{i=1}^{n} \left(y_i - (a x_i^2 + b x_i + c)\right)^2$

obținem poziția optimă a parabolei $y = a x^2 + b x + c$

față de punctele (x_i, y_i), cu $i=\overline{1,n}$.

Rezolvăm sistemul:

$$\begin{cases} \dfrac{\partial E}{\partial a}(a,b,c)=0 \\ \dfrac{\partial E}{\partial b}(a,b,c)=0 \\ \dfrac{\partial E}{\partial c}(a,b,c)=0 \end{cases} \Leftrightarrow \begin{cases} 2\sum\limits_{i=1}^{n}\left(y_i-(ax_i^2+bx_i+c)\right)\cdot(-x_i^2)=0 \,\Big|:2 \\ 2\sum\limits_{i=1}^{n}\left(y_i-(ax_i^2+bx_i+c)\right)\cdot(-x_i)=0 \,\Big|:2 \\ 2\sum\limits_{i=1}^{n}\left(y_i-(ax_i^2+bx_i+c)\right)\cdot(-1)=0 \,\Big|:2 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} a\sum\limits_{i=1}^{n}x_i^4+b\sum\limits_{i=1}^{n}x_i^3+c\sum\limits_{i=1}^{n}x_i^2=\sum\limits_{i=1}^{n}x_i^2y_i \\ a\sum\limits_{i=1}^{n}x_i^3+b\sum\limits_{i=1}^{n}x_i^2+c\sum\limits_{i=1}^{n}x_i=\sum\limits_{i=1}^{n}x_iy_i \\ a\sum\limits_{i=1}^{n}x_i^2+b\sum\limits_{i=1}^{n}x_i+c\underbrace{\sum\limits_{i=1}^{n}1}_{=n}=\sum\limits_{i=1}^{n}y_i. \end{cases}$$

Matricea asociată sistemului este

$$A=\begin{pmatrix} \sum\limits_{i=1}^{n}x_i^4 & \sum\limits_{i=1}^{n}x_i^3 & \sum\limits_{i=1}^{n}x_i^2 \\ \sum\limits_{i=1}^{n}x_i^3 & \sum\limits_{i=1}^{n}x_i^2 & \sum\limits_{i=1}^{n}x_i \\ \sum\limits_{i=1}^{n}x_i^2 & \sum\limits_{i=1}^{n}x_i & \underbrace{\sum\limits_{i=1}^{n}1}_{=n} \end{pmatrix}.$$

Vectorul termenilor liberi este

$$w = \begin{pmatrix} \sum_{i=1}^{n} x_i^2 y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}.$$

3. Regresia exponențială

Dacă datele experimentale descriu o funcție exponențială, atunci curba care le aproximează poate fi aleasă de forma $y = b e^{ax}$.

Considerăm funcția $E(a,b) = \sum_{i=1}^{n} \left(y_i - b e^{a x_i}\right)^2$ și

sistemul
$$\begin{cases} \frac{\partial E}{\partial a}(a,b) = 0 \\ \frac{\partial E}{\partial b}(a,b) = 0 \end{cases} \Leftrightarrow \begin{cases} 2\sum_{i=1}^{n} \left(y_i - b e^{a x_i}\right)\cdot\left(-b e^{a x_i} x_i\right) = 0 \,\Big|:2 \\ 2\sum_{i=1}^{n} \left(y_i - b e^{a x_i}\right)\cdot\left(-e^{a x_i}\right) = 0 \,\Big|:2 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \sum_{i=1}^{n} b^2 e^{2a x_i} x_i - \sum_{i=1}^{n} b e^{a x_i} x_i y_i = 0 \\ \sum_{i=1}^{n} b e^{2a x_i} - \sum_{i=1}^{n} e^{a x_i} y_i = 0. \end{cases}$$

Observăm că acest sistem nu poate fi rezolvat prin metoatele cunoscute.

Pentru a determina a și b logaritmăm expresia

$$y = b e^{ax}.$$

Obținem: $\ln y = \ln b + \ln e^{ax} = \underbrace{\ln b}_{b'} + ax = b' + ax.$

Am notat $b' = \ln b$.

Considerăm $E_1(a, b') = \sum_{i=1}^{n} \left(\ln y_i - (b' + a x_i)\right)^2.$

Rezolvăm sistemul: $\begin{cases} \dfrac{\partial E_1}{\partial a}(a, b') = 0 \\ \dfrac{\partial E_1}{\partial b'}(a, b') = 0 \end{cases} \Leftrightarrow$

$$\Leftrightarrow \begin{cases} 2 \sum_{i=1}^{n} \left(\ln y_i - (b' + a x_i)\right) \cdot (-x_i) = 0 \;\Big|\; :2 \\ 2 \sum_{i=1}^{n} \left(\ln y_i - (b' + a x_i)\right) \cdot (-1) = 0 \;\Big|\; :2 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} a \sum_{i=1}^{n} x_i^2 + b' \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \ln y_i \\ a \sum_{i=1}^{n} x_i + b' \underbrace{\sum_{i=1}^{n} 1}_{n} = \sum_{i=1}^{n} \ln y_i . \end{cases}$$

Matricea asociată acestui sistem este

$$A = \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{n=1}^{n} x_i & \underbrace{\sum_{i=1}^{n} 1}_{= n} \end{pmatrix}.$$

Vectorul termenilor liberi este

$$w = \begin{pmatrix} \sum_{i=1}^{n} x_i \ln y_i \\ \sum_{i=1}^{n} \ln y_i \end{pmatrix}.$$

Rezolvând acest sistem determinăm a și b'. Ținând cont de faptul că $\ln b = b'$ obținem că $b = e^{b'}$.

<u>Definiție</u>. Metoda prezentată în cele trei cazuri de mai sus pentru a determina curbele care aproximează un set de date se numește metoda celor mai mici pătrate.