MINSTAKVADRATMETODEN

INLEDNING (från linjär algebra)
Låt

$$a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n = b_2$$
$$\dots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n = b_m$$

vara ett **olösbart system** (dvs. ett system som saknar lösning).

Vi kan skriva systemet på formen

A**x** = **b** (sys 1)

där

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{och} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

För en vektor **x** kan vi betrakta den så kallade residualvektorn

r (**x**) = A**x** – **b** **(*);**

Uttrycket $|r| = \sqrt{r_1^2 + r_2^2 + \cdots + r_m^2}$ visar hur pass väl en vektor **x** satisfierar systemet.
(Om systemet (sys 1) har en lösning $\mathbf{x_0}$ då är **r** ($\mathbf{x_0}$) = A$\mathbf{x_0}$ – **b** = **0** **)**
Om systemet är olösbart då kan vi bestämma en vektor $\mathbf{x_m}$ som minimerar längden av residualvektorn dvs som minimerar uttrycket

$$|r| = \sqrt{r_1^2 + r_2^2 + \cdots + r_m^2} .$$

Approximationen med den här metoden kallas **minstakvadratmetoden**.

MINSTAKVADRATMETODEN:

Vi multiplicerar (sys 1)
A**x** = **b**
från vänster med transponatet $\mathbf{A^T}$, och får

$\mathbf{A^T A x = A^T b}$ **(sys 2) ,** så kallade **normalsystem.**

Därefter löser vi det nya systemet (sys 2); lösningen till (sys2) minimerar längden av residualvektorn.
Egenskaper för **normalsystemet:**
1. Normalsystemet är kvadratiskt, n ekvationer och n obekanta
2. Normalsystemet (sys2) är alltid lösbart och kan ha exakt en eller oändligt många lösningar.
3. Om normalsystemet har oändligt många lösningar då varje sådan lösning minimerar residualvektorn.

Exempel 1.
Systemet
$x+y=1$
$x+y=2$
$x+2y=2$
$x+2y=3$
är uppenbart olösbart.

i) Bestäm den vektor $\mathbf{x_m}=\begin{bmatrix} x \\ y \end{bmatrix}$ som minimerar residualvektorn.

ii) För detta $\mathbf{x_m}$ bestäm residualvektorn **r(x**

Lösning

a) Vi skriver systemet på matrisform **Ax**=**b**

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix} \qquad (*)$$

b) Systemet (*) multiplicerar vi med A^T och får normalsystem $\mathbf{A^TAx = A^Tb}$ dvs:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix} \Rightarrow$$

$$\begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 8 \\ 13 \end{bmatrix} \qquad \text{(normalsystem på matrisform)}$$

Vi skriver systemet på formen

$$\begin{cases} 4x+6y=8 \\ 6x+10y=13 \end{cases} \quad \mathbf{(**)}$$

och får lösning till (**) x= 1/2 och y=1.

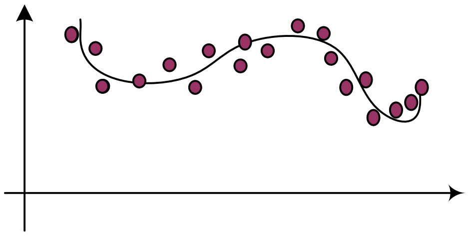
Alltså $\mathbf{x_m}=\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$.

ii) Residualvektorn är $\mathbf{r(x_m)=Ax_m-b}$ = $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}-\begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}=\begin{bmatrix} 3/2 \\ 3/2 \\ 5/2 \\ 5/2 \end{bmatrix}-\begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}=\begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$ och

"felet" = $|\mathbf{r(x_m)}|$= $\sqrt{(1/2)^2+(-1/2)^2+(1/2)^2+(-1/2)^2}=1$.

==

KURVANPASSNING MED MINSTAKVADRATMETODEN.



Vi kan använda minsta kvadrat-metoden (MK-metoden) för att anpassa en kurva $y = f(x)$ med okända koefficienter (t ex $y = ax + b$, $y = ax^2 + bx + c$ eller $y = ae^{bx}$) till experimentdata (mätdata) Y= $(y_1, y_2, \ldots, y_n)$, X= $(x_1, x_2, \ldots, x_n)$
Vi bildar ett ekvationssystem

$$f(x_1) = y_1$$
$$f(x_2) = y_2$$
...
$$f(x_n) = y_n$$

med okända koefficienter a,b,......
Eftersom alla punkter (x_k, y_k) inte ligger (i allmänna fall) på kurvan y=f(x) saknar systemet lösning.
Vi använder MK-metoden och bestämmer koefficienter $a, b, \ldots$ så att kvadratsumman

$$\sum_{k=1}^{n} [y_k - f(x_k)]^2 \qquad \text{(kv.s)}$$

minimeras. (Detta är ekvivalent med att längden av residualvektorn minimeras.)

Kurvan $y = f(x)$ kallas, inom statistiken, **regressionskurvan.**
Typen av kurvan (t ex $y = ax + b$, $y = ax^2 + bx + c$, $y = ae^{bx}$,...) bestämmer vi enligt teoretiska kunskaper om problemet som vi undersöker. Om det saknas teoretisk modell så plottar vi punkterna (x_k, y_k) och därefter väljer vi kurvans typ. Vi kan även testa flera modeller och kolla vilken som gör minsta felet enligt (kv.s).
Exempelvis, i nedanstående Fig 1 med punkterna (x_k, y_k) , kan vi anta att det finns ett linjärt samband $y = ax + b$, mellan X och Y. Sambandet i Fig 2 är uppenbart inte linjärt. Vi kan t. ex. försöka göra kurvanpassning med kurvan $y = ax^2 + bx + c$.

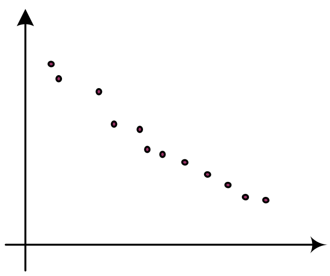


Fig 1

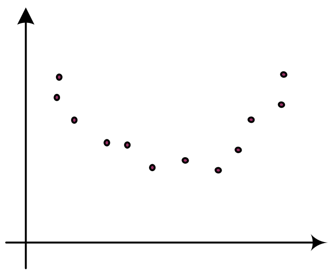


Fig 2

LINJÄRA MODELLER

Anpassning av $y = a + bx$ till mätdata X, Y.

METOD 1 (lin alg)
Vi bildar ett ekvationssystem

$$
\begin{aligned}
ax_1 + b &= y_1 \\
ax_2 + b &= y_2 \\
&\dots \\
ax_n + b &= y_n
\end{aligned}
$$

med okända koefficienter a och b. Därefter skriver vi systemet på matrisform $A\begin{bmatrix} a \\ b \end{bmatrix} = B$

dvs

$$
\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \qquad \text{(sys 1)}
$$

Därefter multiplicerar vi matrisekvationen med A^T och löser systemet
$A^T A = A^T B$ (sys 2)

Vi ser från sys 1 att efter multiplikation med A^T sys 2 kan skrivas på formen

$$
\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix} \qquad \text{(sys 2)}
$$

REGRESSIONSKOEFFICIENT

Regressionskoefficient $r = \dfrac{C(X,Y)}{\sigma_x \sigma_y}$ används som ett mått på hur stark är LINJÄRT samband mellan variablerna.

Uttrycket i täljaren $C(X,Y) = \dfrac{1}{n-1}\left(\sum (x_k - \bar{x})(y_k - \bar{y})\right)$ kallas **kovarians.**

Egenskaper: $-1 \le r \le 1$.
Om punkterna (x_i, y_i) ligger exakt på linjen y=ax+b och a > 0 (resp a<0) då är r=1 (resp r=-1).
Ju närmare r = 1 (eller –1) desto starkare **linjärt** samband mellan X och Y.
Om r är nära 0 då finns det inget linjärt samband mellan X och Y (men då kan finnas ett annat ickelinjärt samband mellan variablerna t ex polynomial, expotential, logaritmiskt eller trigonometriskt samband).

==

Exempel 2. (*Linjär minstakvadratanpassning*)
a) Anpassa linjen $y = ax + b$ enligt minsta kvadrat-metoden till mätdata

X	0	1	2	3
Y	2	1	1	4

och bestäm längden av residualvektorn.
b) Beräkna kovariansen

$$C(X,Y) = \frac{1}{N-1}\left(\sum (x_k - \bar{x})(y_k - \bar{y})\right)$$

(Ekvivalent formel $C(X,Y) = \frac{1}{N-1}\left(\sum x_k y_k - N\overline{xy}\right)$

c) Beräkna korrelationskoefficienten $r = \dfrac{C(X,Y)}{\sigma_x \sigma_y}$.

Är det rimligt att anta att det finns ett linjärt samband mellan s. v, X och Y?

Lösning: (*Lägg märke till att a och b är obekanta.*)

Vi substituerar

X	0	1	2	3
Y	2	1	1	4

i ekvationen $ax + b = y$ och får följande (olösbart) system

$0a + b = 2$
$1a + b = 1$
$2a + b = 1$
$3a + b = 4$

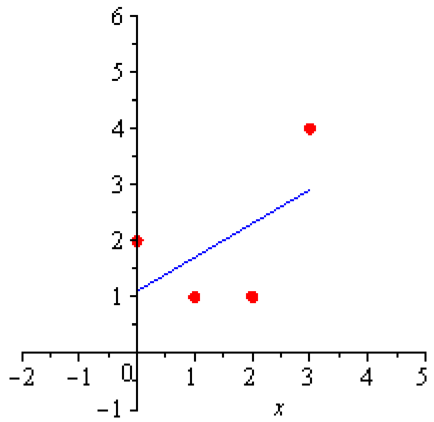
som vi kan skriva på matrisform $\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 4 \end{bmatrix}$

Vi multiplicerar ekvationen från vänster med $A^T = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ och får

$\begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 15 \\ 8 \end{bmatrix}$ som kan skrivas som

$$\begin{cases} 14a + 6b = 15 \\ 6a + 4b = 8 \end{cases} \Rightarrow \begin{cases} 14a + 6b = 15 \\ 3a + 2b = 4 \end{cases} \Rightarrow \begin{cases} 14a + 6b = 15 \\ -9a - 6b = -12 \end{cases} \Rightarrow 5a = 3 \Rightarrow a = 3/5 \Rightarrow b = 11/10$$

Vi ritar punkterna och linjen i samma koordinatsystem:



Svar a: $y = \frac{3}{5}x + \frac{11}{10}$; r= $\begin{bmatrix} -9/10 \\ 7/10 \\ 13/10 \\ -11/10 \end{bmatrix}$ och $|r| \approx 2.05$

b) Svar: $C(X,Y) = \frac{1}{N-1}(\sum x_k y_k - N\overline{x}\overline{y}) = \frac{1}{3}(15 - 4 \bullet 1.5 \bullet 2) = 1$

c) $r = \frac{C(X,Y)}{\sigma_x \sigma_y} = \frac{1}{1.29 \bullet 1.4142} = 0.5477$

Vi kan acceptera antagande om ett linjärt samband om r ligger nära 1 eller -1.
I vårt fall är r=0.5477 långt från 1 och därför kan vi **inte** påstå att det finns **ett linjärt samband** mellan X och Y.

Exempel 3. (*minstakvadratanpassning med en parabel*)
a) Anpassa linjen $y = ax^2 + bx + c$ enligt minstakvadratmetoden till mätdata

X	0	1	2	3
Y	2	1	1	4

och bestäm längden av residualvektorn.

Lösning:
Vi substituerar

X	0	1	2	3
Y	2	1	1	4

i ekvationen $ax^2 + bx + c = y$ och får följande (olösbart) system

$$\begin{aligned} 0a + 0b + c &= 2 \\ 1a + 1b + c &= 1 \\ 4a + 2b + c &= 1 \\ 9a + 3b + c &= 4 \end{aligned}$$

som vi kan skriva på matrisform $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 4 \end{bmatrix}$

Vi multiplicerar ekvationen från vänster med $A^T = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ och får

$$\begin{bmatrix} 98 & 36 & 14 \\ 36 & 14 & 6 \\ 14 & 6 & 4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 41 \\ 15 \\ 8 \end{bmatrix}$$

Härav (efter mycket beräkning)

$a = 1$, $b = -12/5$, $c = 21/10$ och
$y = x^2 - (12/5)x + 21/10$.
Absolutbeloppet av residualvektorn=

$$\left| \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -12/5 \\ 21/10 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 1 \\ 4 \end{bmatrix} \right| = \left| \begin{bmatrix} 1/10 \\ -3/10 \\ 3/10 \\ -1/10 \end{bmatrix} \right| = \sqrt{\frac{20}{100}} = \sqrt{\frac{1}{5}} \approx 0.4472$$

Anmärkning: Felet i den linjära approximationen i exempel 2 är c a 5 gånger större.
Vi ser i nedanstående graf att parabeln, på ett bra sätt, approximerar givna punkter mycket bättre än linjen i exempel 2.

Grafen till punkterna och parabeln $y = x^2 - (12/5)x + 21/10$: