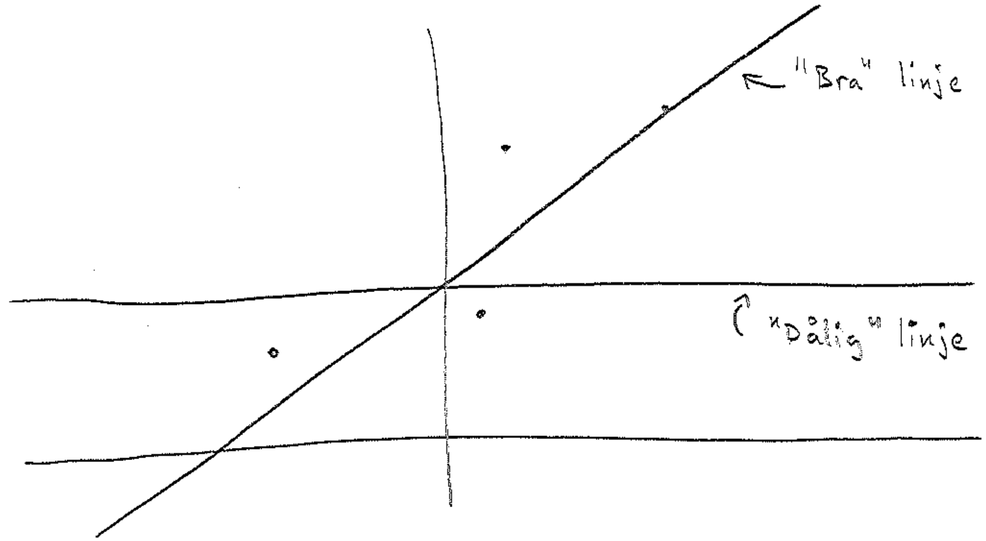
IDAG: Minstakvadrat-metoden och (linjär) regression

Mål: • "Lösa" L.E.S. som saknar lösning

• Approximera experimentell data med t.ex. polynom

<u>Linjär regression</u> har ni förmodligen stött på redan i gymnasiet (utan riktig förklaring).

Det handlar om att givet några par av datapunkter $(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)$ anpassa en rät linje $y = kx + m$ som ligger närmast punkterna.



Vad betyder "närmast"?

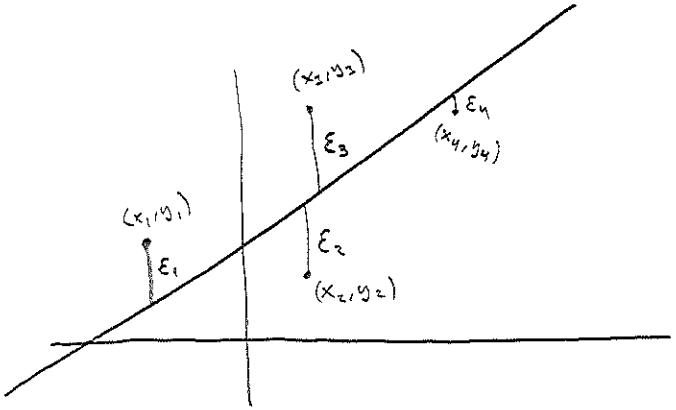
Låt $\varepsilon = (\varepsilon_1, \dots, \varepsilon_p)$ vara vektorn av skillnader,

$$\varepsilon_j = y_j - (kx_j + m) \quad , j = 1, \dots, p.$$

Den bästa approximativa linjen är den som minimerar $\|\varepsilon\|$. Man kan använda olika normer $\| \ \|$, men vi tar den vi definierat, dvs. minimera $\sqrt{\sum_{j=1}^{p} \varepsilon_j^2} = \|\varepsilon\|_2$

OBS: $\iff$ minimera $\sum_{j=1}^{p} \varepsilon_j^2$ (utan kvadratrot)

Visuellt:



Vi kallar y_j <u>observationen</u>,

$kx_j + m$ <u>prediktionen</u> och

ε <u>residualen</u>.

Själva linjen motsvarar en <u>minstakvadrat-lösning</u> (p.g.a. vad vi minimerar).

För att använda samma notation som de flesta ingenjörer och statistiker kommer vi härefter istället för $kx+m$ skriva

$$y = \beta_0 + \beta_1 x$$

β_0 och β_1 är de sökta <u>parametrarna</u>, och vi får p ekvationer:

$$\begin{aligned} \beta_0 + \beta_1 x_1 &= y_1 \\ \beta_0 + \beta_1 x_2 &= y_2 \\ &\vdots \\ \beta_0 + \beta_1 x_p &= y_p . \end{aligned}$$

Detta är uppfyllt om punkterna (x_j, y_j) ligger på linjen, och då är $\varepsilon = 0$.

Vi skriver systemet som $X\beta = y$

där $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_p \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ och $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$.

Om systemet <u>inte</u> har en lösning (dvs. punkterna ligger ej på linjen) så har vi fortfarande

$$\varepsilon = X\beta - y,$$

så vår uppgift är att minimera $\|X\beta - y\|$.

<u>Def.</u> Låt $X \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$. Då är $\hat{\beta}$ en <u>minstakvadrat-lösning</u> till $X\beta = y$ om

$$\|X\hat{\beta} - y\| \leq \|X\beta - y\| \quad \text{för alla } \beta \in \mathbb{R}^n.$$

OBS: Mer generellt X än tidigare! (m punkter/ekv. n variabler)

Eftersom $\{X\beta \mid \beta \in \mathbb{R}^n\} = \mathrm{Col}(X)$ så är $X\hat{\beta}$ den ortogonala projektionen av y på $\mathrm{Col}(X)$.

(Sats 6.3.9)

Vi skriver $\hat{y} = X\hat{\beta} = \mathrm{proj}_{\mathrm{Col}\,X}\, y$.

Vi vet att $\hat{y} - y \in (\mathrm{Col}\,X)^\perp$.

Men enligt Sats 6.1.3 är $(\mathrm{Col}\,X)^\perp = \mathrm{Nul}(X^T)$,

så $X^T(\hat{y} - y) = 0$ dvs. $X^T(X\hat{\beta} - y) = 0$

Alltså har vi

$$\boxed{X^T X \hat{\beta} = X^T y}$$

Detta kallas <u>normalekvationerna</u> för $X\beta = y$.

Notera att om $X \in \mathbb{R}^{m\times n}$ så är $X^T X \in \mathbb{R}^{n\times n}$ och $X^T y \in \mathbb{R}^n$ så om vi har $m > n$ (fler ekv. än variabler) så har ändå normalekv. n ekv. och n variabler.

Ex. Låt $X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$ och $y = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

Då har $X\beta = y$ inga lösningar ($\beta_1 = 0$ och $\beta_1 = 1$!).

Bilda $X^T X = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$

$X^T y = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$

Då ger $X^T X \hat{\beta} = X^T y$ att $\hat{\beta} = \frac{1}{6}\begin{bmatrix} 8 \\ 3 \end{bmatrix}$.

Notera att $X\hat{\beta} = \frac{1}{6}\begin{bmatrix} 5 \\ 8 \\ 11 \end{bmatrix}$ vilket är nära y, men inte lika med y.

Residualen är $\varepsilon = y - \hat{y} = y - X\hat{\beta} = \frac{1}{6}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

Så minstakvadratfelet är $\|\varepsilon\| = \frac{1}{6}\sqrt{1^2 + (-2)^2 + 1^2} = \frac{\sqrt{6}}{6}$

Jämför med $\|y\| = \sqrt{6}$:

Det relativa felet är $\dfrac{\sqrt{6}/6}{\sqrt{6}} = \frac{1}{6} \approx 0.167$,

dvs. 16.7%.

Inte särskilt bra anpassning (men bästa);
kan antyda att punkterna inte följer ett linjärt samband eller att mätfelen är väldigt stora.

I MATLAB fås $\hat{\beta}$ genom $\hat{\beta} = X \backslash y$,

dvs. samma syntax som när $X\beta = y$ har en lösning!

$\longrightarrow$

Därför viktigt att kontrollera om det verkligen är en lösning eller en minstakvadrat-lösning.

MATLAB använder också $A=QR$ istället för normalekvationerna: Läs kort om detta i Lay (och på nästa sida).

Mer generell regression

Istället för linjer kan vi även anpassa polynom av högre grad (vi antog ju bara $X \in \mathbb{R}^{m \times n}$).

Om t.ex. $y = \beta_0 + \beta_1 x + \beta_2 x^2$ får vi

$$X\beta = y \quad \text{där}$$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{och} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Ännu mer generellt kan vi ta

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x).$$

Då blir

$$X = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_m) & f_1(x_m) & \cdots & f_k(x_m) \end{bmatrix}.$$

Minstakvadrat-lösningen minimerar då

$$\|\varepsilon\|^2 = \sum_{j=1}^{m} \left(y_j - \beta_0 f_0(x_j) - \cdots - \beta_k f_k(x_j)\right)^2.$$

MATLAB-DEMO

(Där vi anpassar en linje och ett polynom av grad 3 till data från verkligheten, och ser farorna med extrapolation.)

QR för minsta-kvadrat-problem

Istället för att lösa normalekvationerna direkt kan vi QR-faktorisera först: $X = QR$ med Q ortogonal och R övertriangulär.

Vi får

$$X^T X \beta = X^T y \iff R^T Q^T Q R \beta = R^T Q^T y$$

Men $Q^T Q = I$, och om X har lin. ober. kolonner är R inv. (Lay), så

$$X^T X \beta = X^T y \iff R\beta = Q^T y$$

Triangulärt system: lös med bakåtsubstitution!

Inte billigare beräkning, men säkrare:

Om X är känslig så är $X^T X$ "kvadratiskt" känslig, medan $Q^{-T} R$ är ungefär som X.

Med konditionstal (lab 3):

$$\kappa(X^T X) \approx \kappa(X)^2 \quad \text{men} \quad \kappa(Q^{-T} R) \approx \kappa(X).$$