Anvendt Lineær Algebra

Kursusgang 4

Anita Abildgaard Sillasen

Institut for Matematiske Fag

Vægtet mindste kvadraters metode

For et lineært ligningssystem (af m ligninger med n ubekendte)

$$A\mathbf{x} = \mathbf{b}$$

er en mindste kvadraters løsning $\hat{\mathbf{x}}$ en løsning til *normalligningen*

$$(A^T A)\mathbf{x} = A^T\mathbf{b}.$$

Hvis

$$A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{og} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

så er $A^T A = m$ og $A^T \mathbf{b} = b_1 + \ldots + b_m$ og derfor er

$$\hat{x} = \frac{b_1 + \ldots + b_m}{m},$$

altså (ikke-vægtet) gennemsnit af $b_1, \ldots, b_m$.

Det vægtede gennemsnit af $b_1, \ldots, b_m$ med vægte hhv. $c_1, \ldots, c_m$ (hvor $c_1, \ldots, c_m$ er positive tal) er

$$\frac{c_1 b_1 + \ldots + c_m b_m}{c_1 + \ldots + c_m}.$$

Betragt nu et ligningssystem af m ligninger med n ubekendte

$$A\mathbf{x} = \mathbf{b}.$$

De m ligninger (observationer) tildeles vægte $c_1, \dots, c_m$ som skrives i en matrix

$$C = \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_m \end{bmatrix}.$$

Som vægte kan vi bruge $c_i = \frac{1}{\sigma_i^2}$ hvor σ_i^2 er variansen af den i'te observation, eller eventuelt $c_i = \text{konst.} \frac{1}{\sigma_i^2}$.

Skrives også (i Lay) som $w_i^2 = c_i$ hvor $w_i = \frac{1}{\sigma_i}$ og

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & w_m \end{bmatrix}.$$

Så er $W^T W = WW = C$.

Hvis vægtene $c_1, \ldots, c_m$ er hele tal (positive) så svarer det vægtede ligningssystem til at betragte (ikke-vægtet) ligningssystem

$$\tilde{A}\mathbf{x} = \tilde{\mathbf{b}},$$

hvor ligning nr. i fra ligningssystemet $A\mathbf{x} = \mathbf{b}$ skrives c_i gange.

En mindste kvadraters løsning til det nye ligningssystem er en løsning til normalligningen

$$\tilde{A}^T\tilde{A}\mathbf{x} = \tilde{A}^T\tilde{\mathbf{b}}.$$

Eksempel:

$$A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} r \\ s \\ t \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Så kan vi betragte

$$\tilde{A} = \begin{bmatrix} a & b \\ a & b \\ c & d \\ e & f \\ e & f \\ e & f \end{bmatrix}, \quad \tilde{\mathbf{b}} = \begin{bmatrix} r \\ r \\ s \\ t \\ t \\ t \end{bmatrix}.$$

Normalligningen for ligningssystemet $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$ er

$$\tilde{A}^T \tilde{A}\mathbf{x} = \tilde{A}^T \tilde{\mathbf{b}}.$$

Da $\tilde{A}^T \tilde{A} = A^T CA$ og $\tilde{A}^T \tilde{\mathbf{b}} = A^T C\mathbf{b}$ er denne ligning ækvivalent med

$$A^T CA\mathbf{x} = A^T C\mathbf{b}.$$

Det vægtede ligningssytem $A\mathbf{x} = \mathbf{b}$ med generelle vægte har *normalligning*

$$A^T CA\mathbf{x} = A^T C\mathbf{b}.$$

En *vægtet mindste kvadraters løsning* til $A\mathbf{x} = \mathbf{b}$ er en løsning til normalligningen.

I observationsligningen $A\hat{\mathbf{x}} = \mathbf{b} + \hat{\mathbf{r}}$ indgår residual vektoren

$$\hat{\mathbf{r}} = \begin{bmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_m \end{bmatrix}.$$

Vi udregner

$$\hat{\mathbf{r}}^T C \hat{\mathbf{r}} = [\hat{r}_1 \dots \hat{r}_m] \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_m \end{bmatrix} \begin{bmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_m \end{bmatrix} =$$

$$[\hat{r}_1 \dots \hat{r}_m] \begin{bmatrix} c_1 \hat{r}_1 \\ \vdots \\ c_m \hat{r}_m \end{bmatrix} = c_1 \hat{r}_1^2 + \dots + c_m \hat{r}_m^2.$$

Dette er kvadratet på den vægtede længde af vektoren $\hat{\mathbf{r}}$. Altså kvadratet på den vægtede afstand mellem $A\hat{\mathbf{x}}$ og $\mathbf{b}$.

En vægtet mindste kvadraters løsning er en vektor $\mathbf{x}$ som minimerer

$$\hat{\mathbf{r}}^T C \hat{\mathbf{r}} = (A\mathbf{x} - \mathbf{b})^T C (A\mathbf{x} - \mathbf{b}).$$

Eksempel.

Hvis

$$A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \quad \text{og} \quad C = \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_m \end{bmatrix}$$

så er $A^T CA = c_1 + \ldots + c_m$ og $A^T C\mathbf{b} = c_1 b_1 + \ldots + c_m b_m$ og den vægtede mindste kvadraters løsning er derfor det vægtede gennemsnit

$$\hat{x} = \frac{c_1 b_1 + \ldots + c_m b_m}{c_1 + \ldots + c_m}.$$

Eksempel.

Find linie $y = \alpha x + \beta$ der passer bedst til punkterne

$$(1,3),(2,5),(3,9),(4,12),(5,14)$$

med vægte hhv. 1, 5, 1, 5, 1.

Vi betragter ligningssystemet

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 9 \\ 12 \\ 14 \end{bmatrix}.$$

Normalligningen:

$$(A^T CA)\begin{bmatrix}\beta\\ \alpha\end{bmatrix} = A^T C\mathbf{b}.$$

Den vægtede mindste kvadraters løsning er

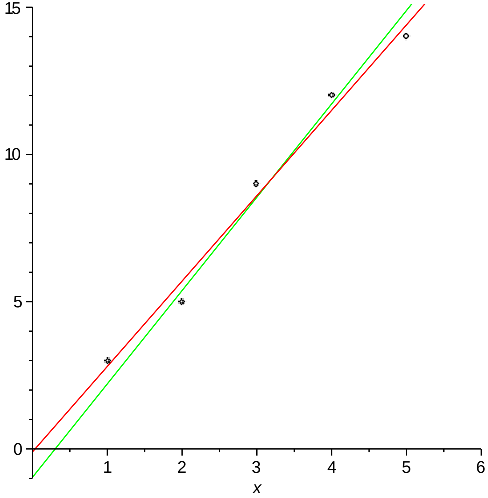
$$\begin{bmatrix}\beta\\ \alpha\end{bmatrix} = (A^T CA)^{-1}A^T C\mathbf{b} = \begin{bmatrix}-0.96\\ 3.17\end{bmatrix}.$$

Den vægtede mindste kvadraters linie har ligning

$$y = 3.17x - 0.96.$$

Denne linie er grøn på næste side.

Den ikke-vægtede mindste kvadraters linie er rød på næste side.


15
10
5
0
1
2
3
4
5
6
x

Den generelle fejlforplantningslov

$X_1, \ldots, X_n$ er stokastiske variable med middelværdier

$$E(X_i)$$

og varianser

$$V(X_i) = E((X_i - E(X_i))^2) = E(X_i^2) - E(X_i)^2.$$

Kovarianserne er

$$\mathrm{Cov}(X_i, X_j) = E((X_i - E(X_i))(X_j - E(X_j))).$$

Hvis X_i og X_j er uafhængige så er $\mathrm{Cov}(X_i, X_j) = 0$.
$\mathrm{Cov}(X_i, X_i) = V(X_i)$.

Kovariansmatricen for

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

er

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \dots & \sigma_n^2 \end{bmatrix},$$

hvor $\sigma_i^2 = V(X_i)$ og
$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i) = \sigma_{ji}$.
$\Sigma_{\mathbf{X}}$ er en symmetrisk matrix.

Hvis en anden vektor af stokastiske variable

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}$$

afhænger lineært af $\mathbf{X}$:

$$\mathbf{Y} = A\mathbf{X}$$

hvor A er en $m \times n$ matrix, så giver fejlforplantningsloven en sammenhæng mellem kovariansmatricerne:

$$\Sigma_{\mathbf{Y}} = A\Sigma_{\mathbf{X}}A^T.$$

En vægtet mindste kvadraters løsning til ligningssystemet $A\mathbf{x} = \mathbf{b}$ med vægtmatrix $C = \operatorname{diag}(c_1, \ldots, c_m)$ afhænger lineært af observationsvektoren $\mathbf{b}$:

$$\hat{\mathbf{x}} = (A^T CA)^{-1} A^T C\mathbf{b}.$$

Vi har valgt C så $c_i = \text{konst.}\frac{1}{\sigma_i^2}$. Dermed er

$$C^{-1} = \begin{bmatrix} c_1^{-1} & 0 & \ldots & 0 \\ 0 & c_2^{-1} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & c_m^{-1} \end{bmatrix} = \frac{1}{\text{konst.}} \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \sigma_m^2 \end{bmatrix} = \frac{1}{\text{konst.}} \Sigma_{\mathbf{b}}.$$

Hvor den sidste lighed skyldes at observationerne antages at være uafhængige. Dermed $\operatorname{Cov}(b_i, b_j) = 0$, når $i \neq j$.

Vi skriver $C = \sigma_0^2 \Sigma_{\mathbf{b}}^{-1}$, hvor σ_0^2 kaldes variansfaktoren.

Variansfaktoren estimeres som

$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{r}}^T C \hat{\mathbf{r}}}{m - n}.$$

Kovariansmatricen $\Sigma_{\hat{\mathbf{x}}}$ kan bestemmes fra fejlforplantningsloven

$$\Sigma_{\hat{\mathbf{x}}} = ((A^T CA)^{-1} A^T C)\Sigma_{\mathbf{b}}((A^T CA)^{-1} A^T C)^T = \sigma_0^2 (A^T CA)^{-1}.$$

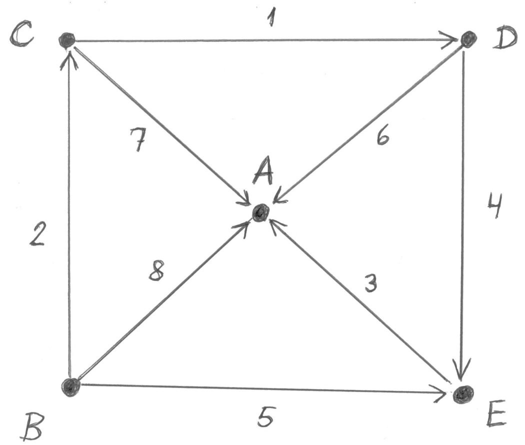
Vi bruger estimatet for σ_0^2:

$$\Sigma_{\hat{\mathbf{x}}} = \hat{\sigma}_0^2 (A^T CA)^{-1}.$$

Fortsættelse af eksempel fra kursusgang 2:
Mellem fem punkter A, B, C, D, E er der målt følgende højdeforskelle:

Fra punkt	Til punkt	Højdeforskel	Afstand
B	A	8	12
B	C	2	20
B	E	5	8
C	A	7	14
C	D	1	16
D	A	6	13
D	E	4	20
E	A	3	8

E er et fikspunkt med højde 10.


1
C
D
7
6
A
4
2
8
3
B
5
E

Sæt

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} , \quad \mathbf{b} = \begin{bmatrix} 8 \\ 2 \\ -5 \\ 7 \\ 1 \\ 6 \\ -6 \\ 13 \end{bmatrix}, \quad \text{og} \quad \hat{\mathbf{r}} = \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \\ \hat{r}_4 \\ \hat{r}_5 \\ \hat{r}_6 \\ \hat{r}_7 \\ \hat{r}_8 \end{bmatrix} .$$

Så er observationsligningen: $A\mathbf{x} = \mathbf{b} + \hat{\mathbf{r}}$.

Som vægtmatrix bruges

$$C = \operatorname{diag}\left(\frac{1}{12}, \frac{1}{20}, \frac{1}{8}, \frac{1}{14}, \frac{1}{16}, \frac{1}{13}, \frac{1}{20}, \frac{1}{8}\right) .$$

En vægtet mindste kvadraters løsning:

$$\hat{\mathbf{x}} = (A^T C A)^{-1} A^T C \mathbf{b} = \begin{bmatrix} 12.9 \\ 4.80 \\ 6.09 \\ 6.73 \end{bmatrix}.$$

Vi får følgende residualvektor

$$\hat{\mathbf{r}} = A\hat{\mathbf{x}} - \mathbf{b} = \begin{bmatrix} 0.117 \\ -0.705 \\ 0.204 \\ -0.178 \\ -0.360 \\ 0.182 \\ -0.730 \\ -0.0878 \end{bmatrix}.$$

Variansfaktor

$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{r}}^T C \hat{\mathbf{r}}}{8-4} = 0.0179.$$

Kovariansmatrix

$$\Sigma_{\hat{\mathbf{x}}} = \hat{\sigma}_0^2 (A^T C A)^{-1} = \begin{bmatrix} 0.083 & 0.039 & 0.061 & 0.054 \\ 0.039 & 0.092 & 0.051 & 0.033 \\ 0.061 & 0.051 & 0.16 & 0.078 \\ 0.054 & 0.033 & 0.078 & 0.14 \end{bmatrix}.$$

Introduktion til Cholesky dekomposition.

Først: repetition af diagonalisering.

En $m \times n$ matrix A er kvadratisk hvis $m = n$.

En kvadratisk matrix A er symmetrisk hvis $A^T = A$.

Idet $(BC)^T = C^T B^T$ er enhver matrix på formen $A^T A$ symmetrisk:
$(A^T A)^T = A^T (A^T)^T = A^T A$.

En vektor $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$ er egenvektor for en $n \times n$ matrix A hvis der findes et tal (en egenværdi) λ så $A\mathbf{x} = \lambda \mathbf{x}$.

En matrix A er diagonaliserbar hvis der findes en basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ for $\mathbb{R}^n$, hvor hver vektor er en egenvektor $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$.
Dette er ækvivalent med at der findes en matrix P og en diagonalmatrix D så $P^{-1}AP = D$.

Enhver symmetrisk matrix er diagonaliserbar.

Enhver matrix på formen $A^T A$ er altså diagonaliserbar.

Hvis λ er en egenværdi for $A^T A$ med en tilhørende egenvektor $\mathbf{x}$, $A^T A\mathbf{x} = \lambda\mathbf{x}$, så er

$$\lambda||\mathbf{x}||^2 = \lambda\mathbf{x}\cdot\mathbf{x} = (A^T A\mathbf{x})\cdot\mathbf{x} = (A^T A\mathbf{x})^T\mathbf{x} =$$

$$\mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = ||A\mathbf{x}||^2 \geq 0.$$

Altså $\lambda \geq 0$.

Hvis søjlerne i A er lineært uafhængige så gælder der faktisk $\lambda > 0$.

En symmetrisk matrix hvis egenværdier alle er > 0 siges at være positiv definit.

En kvadratisk matrix U siges at være en øvre (upper) triangulær matrix hvis der står 0 på alle pladser under diagonalen.

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{bmatrix}.$$

En kvadratisk matrix L siges at være en nedre (lower) triangulær matrix hvis der står 0 på alle pladser over diagonalen.

$$L = \begin{bmatrix} \ell_{11} & 0 & 0 & \cdots & 0 \\ \ell_{21} & \ell_{22} & 0 & \cdots & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{n1} & \ell_{n2} & \ell_{n3} & \cdots & \ell_{nn} \end{bmatrix}.$$

U er en øvre triangulær matrix hvis og kun hvis U^T er nedre triangulær matrix.

En matrix D er en diagonalmatrix hvis og kun hvis den er både øvre triangulær og nedre triangulær.

Cholesky dekomposition

Vi ønsker nu at skrive en $n \times n$ matrix A som et produkt

$$A = U^T U,$$

hvor U er en øvre triangulær matrix.
Dette kaldes Cholesky dekompositionen af A.

I litteraturen skriver man ofte Cholesky dekompositionen som

$$A = LL^T,$$

hvor $L = U^T$ er en nedre triangulær matrix.

MATLAB: chol(A) beregner Cholesky dekomposition U
Maple: LUDecomposition(A,method=Cholesky) beregner $L = U^T$

Enhver matrix på formen $U^T U$ er symmetrisk og hvis tallene på diagonalen i U er $\neq 0$ så er søjlerne lineært uafhængige og dermed er $U^T U$ positiv definit.

Vi vil derfor antage at A er symmetrisk, positiv definit matrix.

Cholesky dekomposition og mindste kvadraters metode

Vi ønsker nu at finde en mindste kvadraters løsning til det (inkonsistente) ligningssystem

$$A\mathbf{x} = \mathbf{b}.$$

Vi betragter derfor normalligningen

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

Matricen $A^T A$ er symmetrisk og da søjlerne i A sædvanligvis er lineært uafhængige er $A^T A$ også positiv definit.

Vi kan derfor finde en Cholesky dekomposition

$$A^T A = U^T U,$$

hvor U er en øvre triangulær matrix og dermed skrive normalligningen som

$$U^T U\mathbf{x} = \mathbf{y},$$

hvor $\mathbf{y} = A^T\mathbf{b}$.

Vi kan løse denne ligning ved at sætte $\mathbf{z} = U\mathbf{x}$ og løse
først

$$U^T\mathbf{z} = \mathbf{y}$$

og derefter

$$U\mathbf{x} = \mathbf{z}.$$

Disse ligninger kan let løses uden brug af rækkeoperationer, da U er triangulær.