6.6. Lineære modeller

Anta at vi har n punkter $(x_1,y_1),\dots,(x_n,y_n)\in \mathbb{R}^2$. Vi vil finne en linje $y=\beta_0+\beta_1 x$ som best approksimerer disse punktene:

$$\sum_{i=1}^{n}\left(y_i -(\beta_0+\beta_1 x_i)\right)^2$$

er minst mulig.

$\{y_i\}_i$ kalles observerte verdier,

$\{\beta_0+\beta_1 x_i\}_i$ kalles predikerte verdier,

$\{y_i-(\beta_0+\beta_1 x_i)\}_i$ kalles residualer.

Dvs vi har et lineært system

$$\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

og vi vil finne en minste kvadraters løsning av dette systemet.

Linje $\beta_0+\beta_1 x$ kalles regresjonslinje, eller minste kvadraters linje; koeffisientene β_0,β_1 kalles regresjonskoeffisientene; $\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$ kalles designmatrisen

Mer generelt, vi kan prøve å bruke andre funksjoner for approksimere (x_i, y_i):

Anta at vi er gitt funksjoner

$$f_1, \ldots, f_m$$

og vi vil bruke lineære kombinasjoner

$$\beta_1 f_1 + \ldots + \beta_m f_m$$

for å approksimer (x_i, y_i) på best mulig måte:

$$\sum_{i=1}^{n} \left(y_i - (\beta_1 f_1(x_i) + \ldots + \beta_m f_m(x_i)) \right)^2$$

er minst mulig.

Dvs. vi vil finne en minste kvadraters løsning av

$$\underbrace{\begin{pmatrix} f_1(x_1) & \ldots & f_m(x_1) \\ \vdots & & \vdots \\ f_1(x_n) & \ldots & f_m(x_n) \end{pmatrix}}_{\substack{X \\ \text{designmatrisen}}} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}}_{\substack{\text{regresjons} \\ \text{koeffisienter}}} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\substack{\text{observerte} \\ \text{verdier}}}$$

Eks.

Betrakt igjen $(0,1), (1,2), (2,1), (3,5) \in \mathbb{R}^2$.

Vi vil finne best approksimasjon til disse i $\mathbb{P}_2$ = kvadratiske polynomer.

$f_1 = 1$, $f_2 = t$, $f_3 = t^2$

$$\text{Designmatrisen } X = \begin{pmatrix} f_1(0) & f_2(0) & f_3(0) \\ f_1(1) & f_2(1) & f_3(1) \\ f_1(2) & f_2(2) & f_3(2) \\ f_1(3) & f_2(3) & f_3(3) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 5 \end{pmatrix}$$

$$\text{Vi har } X^TX = \dots = \begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 35 \\ 14 & 35 & 98 \end{pmatrix},$$

$$X^Ty = \begin{pmatrix} 9 \\ 19 \\ 51 \end{pmatrix}.$$

Vi vil finne en løsning av normallikningen $X^TX\beta = X^Ty$. Vi får en løsning

$$\beta = \begin{pmatrix} 27/20 \\ -23/20 \\ 3/4 \end{pmatrix}. \qquad \boxed{\frac{27}{20} - \frac{23}{20}t + \frac{3}{4}t^2}$$

7.1 Symmetriske matriser

Def

En $n \times n$ matrise $A = (a_{ij})_{i,j=1}^{n}$ kalles **symmetrisk** hvis

$$A^T = A \quad , \text{ dvs. } a_{ij} = a_{ji} \text{ for alle } i,j.$$

Eks.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{ikke symmetrisk}$$

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \quad \text{symmetrisk}$$

Lemma

Anta at $A = (a_{ij})_{i,j=1}^{n}$ er en reel matrise.

Da er A symmetrisk hvis og bare hvis

$$\langle Ax, y \rangle = \langle x, Ay \rangle \quad \text{for alle } x, y \in \mathbb{R}^n.$$

($\langle \cdot, \cdot \rangle$ er prikkproduktet: $\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i$)

Bevis

$$\langle Ax, y \rangle = \left\langle \begin{pmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \cdots \\ \sum_{j=1}^{n} a_{nj} x_j \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{i,j=1}^{n} a_{ij} x_j y_i ,$$

$$\langle x, By \rangle = \sum_{i,j=1}^{n} b_{ij} x_i y_j = \sum_{i,j=1}^{n} b_{ji} x_j y_i \quad \text{for } n \times n \text{ matrise } B$$

Det følger at $\langle Ax, y \rangle = \langle x, A^T y \rangle$ for alle A.

(Eller: $\langle Ax, y \rangle = (Ax) \cdot y = (Ax)^T y = x^T A^T y = \langle x, A^T y \rangle$.)

Anta nå at A er en symmetrisk matrise.
Vi har

$$\langle Ax, y\rangle = \langle x, A^T y\rangle = \langle x, Ay\rangle.$$

Omvendt, anta at $\langle Ax, y\rangle = \langle x, Ay\rangle$ for alle x, y.
Ta $x = e_i$, $y = e_j$ $\left(e_i = (0, \dots, 0, \underset{i}{1}, 0, \dots, 0) \right)$
Da har vi

$$\langle Ae_i, e_j\rangle = a_{ji}, \quad \langle e_i, Ae_j\rangle = a_{ij}.$$

Vi ser at $a_{ij} = a_{ji}$. ∎

<u>To viktigste egenskaper av symmetriske reelle matriser:</u>

Anta at A er en reel symmetrisk matrise. Da har vi:

1) A har bare reelle egenverdier;
2) (Teorem 1) Hvis $v_1, v_2 \in \mathbb{R}^n$ er egenvektorer til A, $Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2$, $\lambda_1 \neq \lambda_2$, da

$$v_1 \perp v_2.$$

<u>Bevis</u>

1) Anta at $\lambda \in \mathbb{C}$ er en rot av p_A = karakteristiske polynomen til A. Da har vi en egenvektor $v \in \mathbb{C}^n$: $Av = \lambda v$, $v \neq 0$.

Betrakt $c = \sum_{i,j=1}^n a_{ij} v_j \bar{v}_i \in \mathbb{C}$.

Vi har

$$c = \sum_{i=1}^n \Big(\sum_{j=1}^n a_{ij} v_j \Big) \bar{v}_i = \underline{\sum_{i=1}^n \lambda v_i \bar{v}_i} = \lambda \sum_{i=1}^n |v_i|^2,$$

$$c = \sum_{j=1}^n \Big(\sum_{i=1}^n a_{ij} \bar{v}_i \Big) v_j = \sum_{j=1}^n \overline{\Big(\sum_{i=1}^n a_{ji} v_i \Big)} v_j = \sum_{j=1}^n \overline{\lambda v_j}\, v_j = \bar{\lambda} \sum_{i=1}^n |v_i|^2$$

Vi ser at
$$\lambda \sum_{i=1}^{n} |v_i|^2 = \overline{\lambda} \sum_{i=1}^{n} |v_i|^2.$$
Vi konkluderer at $\lambda = \overline{\lambda}$, så $\lambda \in \mathbb{R}$.

2) $Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2$, $\lambda_1 \neq \lambda_2$.
Betrakte $c = \langle Av_1, v_2 \rangle$.
Vi har
$$c = \langle Av_1, v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle,$$
$$c = \langle Av_1, v_2 \rangle = \langle v_1, Av_2 \rangle = \langle v_1, \lambda_2 v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle.$$
Vi ser $\lambda_1 \langle v_1, v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$, dvs.
$$(\lambda_1 - \lambda_2) \langle v_1, v_2 \rangle = 0.$$
Vi konkluderer at $\langle v_1, v_2 \rangle = 0$. ∎

Merk: 1) er ikke sant for symmetriske komplekse matriser. Men 1) er sant for komplekse selv-adjungerte, eller Hermetiske, matriser, som er definert ved
$$a_{ij} = \overline{a_{ji}}.$$

Def
En $n \times n$ reel matrise A er ortogonalt diagonaliserbar hvis det finnes en ortogonal basis i $\mathbb{R}^n$ som består av egenvektorer til A.
Dvs., det finnes en ortogonal matrise P ($P^{-1} = P^T$) og en diagonal matrise D slik at
$$A = PDP^{-1} = PDP^T.$$

Teorem 2

En $n \times n$ reel matrise A er ortogonalt diagonaliserbar hvis og bare hvis A er symmetrisk.

Bevis

Anta at A er ortogonalt diagonaliserbar:

$$A = PDP^T$$

for en ortogonal matrise P og en diagonalmatrise D.

Da $A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$.

Anta nå at A er en symmetrisk reel matrise.
La $\lambda_1, \dots, \lambda_k$ være forskjellige egenverdier til A.

Betrakt $V_i = \{v \in \mathbb{R}^n : Av = \lambda_i v\}$, $i = 1, \dots, k$.
Vi vet at $V_i \perp V_j$ for $i \neq j$.

La $\mathcal{B}_i$ være en ortonormal basis for V_i.
Da er $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$ er et ortonormalt system som består av egenvektorer til A.
Vi vil vise at $\mathcal{B}$ er en basis.
Vi trenger å vise at $\operatorname{span} \mathcal{B} = \mathbb{R}^n$.
Dette er det samme som å vise at $W = (\operatorname{span} \mathcal{B})^\perp = 0$.