

Tras dat - kovariančna matrica

kovariančna matrica : $S = \frac{1}{n-1} \sum_{i=1}^n \underbrace{(x_i - \bar{x})}_{d \times 1} \underbrace{(x_i - \bar{x})^T}_{1 \times d} =$

data: $X = \begin{bmatrix} -x_1^T \\ \vdots \\ -x_n^T \end{bmatrix}_{n \times d} = \begin{bmatrix} otec_1 & syn_1 \\ \vdots & \vdots \\ otec_n & syn_n \end{bmatrix}_{n \times 2}$

$$\begin{bmatrix} \hat{\sigma}_1^2 & \dots & \hat{cov}_{1,d} \\ \vdots & \ddots & \vdots \\ \hat{cov}_{d,1} & \dots & \hat{\sigma}_d^2 \end{bmatrix}_{d \times d}$$

$\bar{x} = \begin{bmatrix} \frac{1}{n} \sum otec_i \\ \frac{1}{n} \sum syn_i \end{bmatrix} = \frac{1}{n} \begin{bmatrix} otec_1 & \dots & otec_n \\ syn_1 & \dots & syn_n \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{n} X^T \mathbf{1}_n$
 $\uparrow$ $\mathbf{1}_n$ je vektor.

$S = \frac{1}{n-1} \begin{bmatrix} | & & | \\ x_1 - \bar{x} & \dots & x_n - \bar{x} \\ | & & | \end{bmatrix}_{d \times n} \begin{bmatrix} -(x_1 - \bar{x})^T \\ \vdots \\ -(x_n - \bar{x})^T \end{bmatrix}_{n \times d} = \frac{1}{n-1} B B^T$

$\Rightarrow S$: sym.
: h.s.d.

$(x^T B B^T x = y^T y = \|y\|^2 \geq 0 \quad \forall x)$

$= \dots = \frac{1}{n-1} X^T \underbrace{\left(I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right)}_P \underbrace{\left(I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right)^T}_{P^T} X$

P : projekcijska matrica
: projektuje na $\mathbf{1}_n^\perp$

korelacijska matrica $R = \begin{bmatrix} 1 & \dots & r_{1,d} \\ \vdots & \ddots & \vdots \\ r_{d,1} & \dots & 1 \end{bmatrix}$

$r_{i,j} = \frac{\hat{cov}_{ij}}{\sqrt{\hat{\sigma}_i \hat{\sigma}_j}} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}} \in [-1, 1]$

$R = \begin{bmatrix} 1/\sqrt{s_{11}} & 0 \\ 0 & 1/\sqrt{s_{dd}} \end{bmatrix} S \begin{bmatrix} 1/\sqrt{s_{11}} & 0 \\ 0 & 1/\sqrt{s_{dd}} \end{bmatrix}$

Zachytit tras dat $\rightarrow$ Norm. rozdelenie

d -perm. - d majú Norm. rozdelenie (d -normálne) $\rightarrow$ so str. hodnotami $\mu \in \mathbb{R}^d$
 $\rightarrow$ s kovar. maticou $\Sigma \in \mathbb{R}^{d \times d}$

hustota $f(x) \dots$ vektormice : $f(x) = c$
 $x \in \mathbb{R}^d$

$(x - \mu)^T \Sigma^{-1} (x - \mu) = c$
 kvadratická forma

$\dots$ smerý natiálnutia (hl. polosi)
 $\rightarrow$ vektorý Σ
 $\rightarrow \Sigma = Q \Lambda Q^T$

Zachytit tras dat : PCA

PCA - Principal component analysis

Analýza (Metóda) hlavných komponentov

- 2-normálna reprezentácia d -normálnych dat



Analýza metód / hlavných komponentov

- 2-rozmerná reprezentácia 2-rozmerných dát

S : (vy'herová) kov. mat.

$$S = Q \Lambda Q^T \rightarrow \begin{cases} q_1: \text{smerník najväčšieho rozptylu} \\ q_2: \text{2. najv.} \\ q_3: \text{3. najv.} \\ \vdots \\ q_n: \text{najmenšieho rozptylu dát} \end{cases} \quad \begin{matrix} \text{(zoradené} \\ \text{vl. hodnoty: } \lambda_1 \geq \lambda_2 \geq \dots \\ \geq \lambda_n) \end{matrix}$$

$\Rightarrow$ vezme sme iba prvých (najv.) hlavných smerov:

0, vycentrovane dát: aby stred bol v 0_n

1, vypočítat S

2, $S = Q \Lambda Q^T$

3, zoberieme iba q_1, q_2

4, reprezentácia x -ov pomocou q_1, q_2

$\downarrow$
prídame súradnice x_i vzhľadom na
súradnicovú sústavu danú $q_1, \dots, q_n$

súradnice x_i vo svete $q_1, \dots, q_n$: $y_i \in \mathbb{R}^2$

$$x_i = y_i^{(1)} q_1 + \dots + y_i^{(n)} q_n = \begin{bmatrix} q_1 & \dots & q_n \end{bmatrix} \begin{bmatrix} y_i^{(1)} \\ \vdots \\ y_i^{(n)} \end{bmatrix}$$

$$x_i = Q y_i \quad / \quad Q^T (Q^T Q = I)$$

$y_i = Q^T x_i$ $\rightarrow$ vezme sme prvých 2 súradnice tohto vektora

$$\begin{bmatrix} y_1 & \dots & y_n \end{bmatrix} = Q^T \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$$

$$Y^T_{n \times m} = Q^T X^T_{m \times n}$$

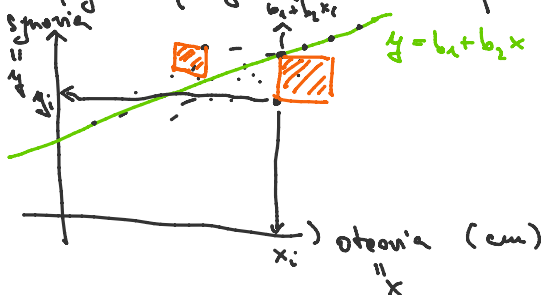
$\rightarrow$ zoberat prvých 2 stĺpce

$$Y = X Q$$

• nielen 2-D reprezentácia, ale napr. aj 3D-rozm. reprezentácia 720-rozmerných dát

Lineárna regresia

• vplyv nejakých prem. na nejakú inú



: máme $x_1, \dots, x_n \in \mathbb{R}$ (otvor)

$y_1, \dots, y_n \in \mathbb{R}$ (synovia)

$\rightarrow$ vzťah medzi x a y : $y = b_1 + b_2 x$ $b_1, b_2 = ?$

lineárny Ideaľ: všetky (x_i, y_i) na priamke:

$$y_i = b_1 + b_2 x_i$$

$$y_1 = b_1 + b_2 x_1$$

$\vdots$

$$y_n = b_1 + b_2 x_n$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$y = X b$$

$$b = A^{-1} X$$

• všetky na
priamke
rešenie

nie je konzistentný

$A^T A x = A^T b \Rightarrow$ aproximácia : $\| X b - y \|^2 \rightarrow \min_b$: metóda
najmenších
štvorcov
MNS

→ norm. rovnice : $X^T X b = X^T y$

⇒ rieš. : $b = (X^T X)^{-1} X^T y$
 $= (X^T X)^{-1} X^T y$ ak $\text{rank}(X) = 2$

n zmeraní statistikov : $y = X \beta + \varepsilon$

→ odhad : MNS : $\hat{\beta} = (X^T X)^{-1} X^T y$

• môžeme mať aj viac premenných : $y_i = b_1 + b_2 x_i + b_3 t_i + b_4 w_i + \dots + b_q p_i$

⇓

$$X : n \times q$$

→ vplyvy jednotlivých premenných na y .