RDD - 05

15 March 2021 14:41

- $I(z)$ z minula - overené

Projekčné sledovanie (PP)

- dynamické vykresľovanie: tours - výlety

animate_xy (... guided_tour ..,) → PP

(... grand_tour ...) → všetky možné pohľady na dáta

Faktorová analýza (Factor analysis, FA)

- hľadá skryté faktory → špeciálne: rozvoj teórie

→ jednoduchšie vyjadrenie dát / ukazovateľov

- psychológia, sociológia, teraz aj marketing, medicína ...

teoretický: náh. vektor $\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ - popísať pomocou skrytých (latentných) faktorov $F_1, \dots, F_q$, $q < k$

model: náh. prem.

$$X_i = \sum_{j=1}^{q} a_{ij} F_j + U_i \quad , \quad i = 1, \dots, k$$

faktorové náklady (factor loadings)

skryté (latentné) faktory

špecifické faktory

maticovo: $\mathbb{X} = A\mathbb{F} + \mathbb{U}$

$A = \begin{bmatrix} a_{11} & \dots & a_{1q} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kq} \end{bmatrix}$ - neznáme parametre modelu

$\mathbb{F} = \begin{pmatrix} F_1 \\ \vdots \\ F_q \end{pmatrix}$, $\mathbb{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_k \end{pmatrix}$

uniqueness = vekt. špecif. faktorov

(odteraz: X, U, F = $\mathbb{X}, \mathbb{U}, \mathbb{F}$)

Predpoklady:

$Cov(F, U) = 0_{q \times k}$

$Var(F) = I_q$, $E(F) = 0_q$

$Var(U) = D = diag(\underbrace{d_1, \dots, d_k}_{\text{tiež neznáme parametre modelu}})$

$E(U) = 0_k$

pre jednoduchosť: $E(X_i) = 0$, $D(X_i) = 1$ $\forall i = 1, \dots, k$

$$\Rightarrow \underbrace{Var(\mathbb{X})}_{=\Sigma} = Var(A\mathbb{F} + \mathbb{U}) = Var(AF) + Var(U) = A \underbrace{Var(F)}_{=I} A^T + \underbrace{Var(U)}_{=D} = AA^T + D$$

$cov = 0$

$$\boxed{\Sigma = AA^T + D}$$

- $Cov(X, F) = Cov(AF + U, F) = Cov(AF, F) + Cov(U, F) = A \cdot I_q = A$

$$\mathrm{Cov}(X,F) = \mathrm{Cov}(AF+U,F) = \underset{A\cdot \mathrm{Cov}(F,F)}{\underset{\parallel}{\mathrm{Cov}(AF,F)}} + \underset{0}{\underbrace{\mathrm{Cov}(U,F)}} = A\cdot I_q = A$$

$$\Rightarrow \rho(X_i,F_j) = \frac{a_{ij}}{\sqrt{\underset{1}{\underset{\parallel}{D(X_i)}}\,\underset{1}{\underset{\parallel}{D(F_j)}}}} = a_{ij} \quad //$$

• vezmime rotáciu A, F: $\tilde{A} = AV$, $\tilde{F} = V^T F$, V – ortog. matica

$$X = AF + U$$

$$\tilde{A}\tilde{F} + U = A\underbrace{VV^T}_{I}F + U = AF + U = X$$

$\Rightarrow$ <u>preparametrizovaný model</u>: ľub. rotácia matice parametrov A tiež vyhovuje modelu

$\rightarrow$ nejednoznačnosť matice fakt. nákladov A

$\Rightarrow$ veľa rovnako dobrých modelov daných AV, V-ortog. Môžem si vybrať ten, ktorý mi vyhovuje.

<u>prakticky</u>: 3 kroky FA:

<u>1.) Odhad A</u> (fakt. nákladov)

$$\underset{\mathrm{Var}(X)}{\underset{\parallel}{\Sigma}} = AA^T + D$$

a) <u>metóda hlavných faktorov</u>: (m. of principal factors)

$$\underset{\mathrm{Var}(X) = \begin{pmatrix} 1 & & * \\ & \ddots & \\ * & & 1 \end{pmatrix}}{\underset{\parallel}{\Sigma}} = AA^T + D = \begin{bmatrix} \underbrace{\sum_{j=1}^{q} a_{1j}^2}_{=h_1^2} & & * \\ & \ddots & \\ * & & \underbrace{\sum_{j=1}^{q} a_{kj}^2}_{=h_k^2} \end{bmatrix} + \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_q \end{bmatrix}$$

– odhadneme ju ako R (korel. matica)

: komunality

$\hookrightarrow R_{ij} = \dfrac{S_{ij}}{\sqrt{S_{ii}\cdot S_{jj}}}$ $\leftarrow$ kovar. matica

1. Odhadnúť h_i a d_i

$$\hat{h}_i^2 = 1 - \frac{1}{R^{-1}_{ii}} \quad \Rightarrow \quad 1 = h_i^2 + d_i$$

$$\rightarrow \hat{d}_i = 1 - \hat{h}_i^2 = \frac{1}{R^{-1}_{ii}}$$

2. odhadnúť AA^T: $R = AA^T + D$

$$\Rightarrow \widehat{AA^T} = R - \hat{D} = R^* = R - \mathrm{diag}\left(\frac{1}{R^{-1}_{11}}, \dots, \frac{1}{R^{-1}_{kk}}\right)$$

3. odhadnúť A

$$\underset{\substack{\uparrow \\ \text{spektr. rozklad}}}{AA^T} = W \Lambda W = \underbrace{W \Lambda^{1/2}}_{k \times k} \underbrace{\Lambda^{1/2} W^T}_{k \times k}$$ my chceme $A : k \times q$

$\lambda_1 \geq \ldots \geq \lambda_k$. ak prvých q λ_i sú ≥ 0 a dosť veľké $\Rightarrow$

$$AA^T \approx \underbrace{\underbrace{\begin{pmatrix} | & & | \\ w_1 & \cdots & w_q \\ | & & | \end{pmatrix}}_{W_q} \underbrace{\begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_q} \end{pmatrix}}_{\Lambda_q^{1/2}}}_{\hat{A}} \underbrace{\begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_q} \end{pmatrix} \begin{pmatrix} - w_1^T - \\ \vdots \\ - w_q^T - \end{pmatrix}}_{\hat{A}^T}$$

$$\Rightarrow \hat{A} = W_q \Lambda_q^{1/2}$$

... môžeme pokračovať : nové $\hat{h}_i^2 = \sum_{j=1}^{q} \hat{a}_{ij}^2$ a opakovať proces

$\rightarrow$ iterovať, až kým sa komunality neustália.

b) met. max. vierohodnosti:

predp.: $x_1, \ldots, x_n$: náh. výber z $N_k(0_k, AA^T + D)$, $(AA^T + D)_{ii} = 1 \ \forall i$

$$\Rightarrow L(A, D) = \prod_{i=1}^{n} f_{N_k(0_k, AA^T+D)}(x_i) \rightarrow \max$$

A; D-diag.

$A^T D^{-1} A$ -diag. $\leftarrow$ bez toho by bolo ∞ – veľa rieš.; vždy také rieš. existuje

2.) Rotácia faktorov

- hľadáme V – ortog., aby $\hat{A}V$ bolo "pekné"

$\rightarrow$ aby bolo dobre interpretovateľné.

$a_{ij} = \rho(X_i, F_j)$: chceme $\tilde{a}_{ij}$ blízke $\begin{cases} 1 & : \text{veľká korel.} \\ 0 & : \text{nekorelované} \\ -1 & : \text{veľká neg. korel.} \end{cases}$

nech $B_V = \hat{A}V \underset{\substack{\uparrow \\ \text{Hadamardov súčin}}}{\odot} \hat{A}V$ $\quad (B_V)_{ij} = (\hat{A}V)_{ij}^2$

$\Rightarrow$ chceme $(B_V)_{ij}$ blízke $\begin{cases} 1 \\ 0 \end{cases}$

$(\hat{A}V)_{ij}$ $\begin{cases} 1 \\ 0 \\ -1 \end{cases}$ $\Rightarrow$

$$B_V = \begin{bmatrix} b_{11} & \cdots & b_{1q} \\ \vdots & & \\ b_{k1} & \cdots & b_{kq} \end{bmatrix}$$

$\max f(b_{11} \ldots b_{kq})$, aby to vyplnulo $b_{ij} \approx \begin{cases} 1 \\ 0 \end{cases}$

$\max f(b_{11} \dots b_{kq})$, aby to vyzeralo $b_{ij} \doteq \begin{cases} 0 \\ 1 \end{cases}$

↳ disperzia, lebo tá je maximálna, ak ½ čísel sú 0 a ½ sú 1.

Možnosti: a) výberová disperzia:

$$f_{kv.}(V) = \frac{1}{kq} \sum_{i=1}^{k} \sum_{j=1}^{q} (b_{ij} - \overline{B}_V)^2$$

priemerov všetkých b_{ij}

$$= \text{—''—} \left((\hat{A}V)_{ij}^2 - \overline{(\hat{A}V) \odot \hat{A}V} \right)$$

dá sa ukázať: $f_{kv.}(V) = \frac{1}{kq} \sum_{i=1}^{k} \sum_{j=1}^{q} (\hat{A}V)_{ij}^{4} - \frac{1}{kq} \| \hat{A} \|^{4}$ → kvartimaxná metóda

kvarti-

po prvkoch

$$V^* \in \underset{V\text{-ortog.}}{\operatorname{argmax}} f_{kv}(V)$$

b) súčet disperzií po stĺpcoch

$$B = \begin{bmatrix} b_{11} & \dots & b_{1q} \\ \vdots & & \\ b_{k1} & \dots & b_{kq} \end{bmatrix}$$

$\downarrow \qquad \downarrow$

$\sum D() \quad D() \to \max$

$$\frac{1}{kq} \sum_{j=1}^{q} \sum_{i=1}^{k} (b_{ij} - \overline{b}_j)^2 =: f_{var}(V)$$: varimaxná metóda

$$\frac{1}{kq} \sum_i \sum_j \left((\hat{A}V)_{ij}^2 - \frac{1}{k} \sum_{s=1}^{k} (\hat{A}V)_{sj}^2 \right)^2 = ?$$

$$V^* = \underset{V\text{-ort.}}{\operatorname{argmax}} f_{var}(V)$$

3.) Odhady faktorových skóre

- i-ty objekt: $x_i \in \mathbb{R}^k$ – namerané $(i = 1, \dots, n)$

  : $F^{(i)} \in \mathbb{R}^q$ – vektor (skutočných) skrytých faktorov pre i-ty objekt

  : odhad: $\hat{F}^{(i)}$ – odhad faktorového skóre pre i-ty objekt

$x_1, \dots, x_n \in \mathbb{R}^k$

$\downarrow \qquad \downarrow$

$F^{(1)} \quad F^{(n)} \in \mathbb{R}^q$

$$\mathbb{F}^{(1)} \; \dots \; \mathbb{F}^{(n)} \in \mathbb{R}^q$$

$$\operatorname{Var}\left(\begin{pmatrix} \mathbb{X} \\ \mathbb{F} \end{pmatrix}\right) = \operatorname{Var}\left(\begin{pmatrix} A\mathbb{F} + \mathbb{U} \\ \mathbb{F} \end{pmatrix}\right) = \operatorname{Var}\left(\underbrace{\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}}\begin{pmatrix} \mathbb{U} \\ \mathbb{F} \end{pmatrix}\right) = \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \underbrace{\operatorname{Var}\begin{pmatrix} U \\ F \end{pmatrix}}_{= \begin{pmatrix} D & 0 \\ 0 & I \end{pmatrix}} \begin{pmatrix} I & 0 \\ A^T & I \end{pmatrix}$$

$$= \begin{pmatrix} AA^T + D & A \\ A^T & I \end{pmatrix}$$

Predp.: $\begin{pmatrix} \mathbb{X} \\ \mathbb{F} \end{pmatrix} \sim N_{k+q}$

Lema: Nech $\begin{pmatrix} X \\ F \end{pmatrix} \sim N_{k+q}\left(0, \begin{pmatrix} AA^T + D & A \\ A^T & I \end{pmatrix}\right)$. Potom podmienené rozdelenie F za podmienky $X = x$ je $F \mid X = x \sim N_q\left(A^T (AA^T + D)^{-1} x, \text{ čosi}\right)$.

$\Rightarrow$ ak i-ty objekt má $X = x_i$, potom $E(F \mid X = x_i) = A^T (AA^T + D)^{-1} x_i$

$\Rightarrow$ odhadneme $\hat{F}^{(i)} = A^T (AA^T + D)^{-1} x_i$, $i = 1, \dots, n$.