RDD - 08

13 April 2021 15:15

Mnohorozmerné škálovanie (MŠ) – multidimensional scaling (distance scaling)

- máme maticu vzájom. vzdialeností: $D \in \mathbb{R}^{n \times n}$, D_{ij} = vzdialenosť medzi i a j
- nemáme pôvodné $x_1, \dots, x_n \in \mathbb{R}^q$

→ vykresliť v 2/3-rozmere, reprezentovať v q-rozmere

⇒ hľadáme $y_1, \dots, y_n \in \mathbb{R}^q$: $\|y_i - y_j\| \approx D_{ij}$

- ozaj vzdialenosti (mestá, ...), nepodobnosti (rozdielne zloženie výrobkov, ...)

1.) Klasické MŠ

- posunutie všet. bodov nemení vzdialenosti ⇒ BUNV: $\sum_{i=1}^n x_i = 0_q$ (centrované)

$$\underset{\text{Gramova matica}}{K} = XX^T = \begin{bmatrix} - x_1^T - \\ \vdots \\ - x_n^T - \end{bmatrix} \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix} = \begin{pmatrix} x_1^T x_1 & \dots & x_1^T x_n \\ \vdots & & \vdots \\ x_n^T x_1 & \dots & x_n^T x_n \end{pmatrix} \quad : \quad K_{ij} = \langle x_i, x_j \rangle = x_i^T x_j$$

- matica skal. súčinov

Lema (od Gram. matice ku vzdialenostiam):

$K = XX^T$, $\sum_{i=1}^n x_i = 0_q$

$$\Rightarrow \|x_i - x_j\|^2 = K_{ii} + K_{jj} - 2K_{ij}.$$

→ od K ku D.

Naopak: od D ku K:

Youngova – Hausholderova veta: [$\sum x_i = 0$ stále]

$D_{ij} = \|x_i - x_j\|$ ⇒ pre $K = XX^T$ platí:

$$K_{ij} = -\frac{1}{2}\left[D_{ij}^2 - \frac{1}{n}\sum_{l=1}^n D_{il}^2 - \frac{1}{n}\sum_{p=1}^n D_{pj}^2 + \frac{1}{n^2}\sum_{p=1}^n \sum_{l=1}^n D_{pl}^2 \right] \quad \forall i,j.$$

→ Bez dôkazu

Maticovo: $A \in \mathbb{R}^{n \times n}$: $A_{ij} = -\frac{1}{2} D_{ij}^2$, $P = I_n - \frac{1}{n} J_n \in \mathbb{R}^{n \times n}$

$$\Rightarrow K = PAP.$$

$K = XX^T$ (X – nepoznám) ⇒ K – k.s.d.

↑ vypočítané ⇒ spektr. rozklad: $K = \overset{n\times n}{U} \overset{n\times n}{\Lambda} \overset{n\times n}{U^T}$

↖ vypočítané'

$\Rightarrow$ spekt. rozklad : $K = U \Lambda U^T$
$\uparrow$ orthog. $\uparrow$ diag. : diag$(\lambda_1, \dots, \lambda_n)$

• ak rank $(K) = r$ $\Rightarrow$ $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$

$$\rightarrow K = \underbrace{\begin{pmatrix} u_1 \dots u_r \end{pmatrix}}_{\substack{U_1 \\ n \times r}} \underbrace{\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix}}_{\Lambda_1 \; r\times r} \underbrace{\begin{pmatrix} u_1^T \\ \vdots \\ u_r^T \end{pmatrix}}_{r \times n} = \underset{n\times r}{U_1} \underset{r \times r}{\Lambda_1} U_1^T$$

$$= \underbrace{U_1 \Lambda_1^{1/2}}_{Y \in \mathbb{R}^{n \times r}} \underbrace{\Lambda_1^{1/2} U_1^T}_{Y^T}$$

$\rightarrow$ $K = YY^T$

$$Y = \begin{pmatrix} - y_1^T - \\ \vdots \\ - y_n^T - \end{pmatrix}, \quad y_i \in \mathbb{R}^r$$

$\rightarrow$ presne reprezentujú vzdialenosti : $\|y_i - y_j\| = D_{ij}$

• menej rozmerov : BUNV: $\lambda_1 \geq \dots \geq \lambda_r > 0$

$\rightarrow$ vezmeme prvých q vl. hodnôt a vektorov:

$$Y = \underset{n \times q}{\begin{pmatrix} | & & | \\ u_1, & \dots & u_q \\ | & & | \end{pmatrix}} \underset{q \times q}{\begin{pmatrix} \lambda_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_q^{1/2} \end{pmatrix}} \in \mathbb{R}^{n \times q} \quad : \quad Y = \begin{pmatrix} y_1^T \\ \vdots \\ y_n^T \end{pmatrix} : y_i \in \mathbb{R}^q$$

$\rightarrow$ potom už iba aproximujeme: $\|y_i - y_j\| \approx D_{ij}$

<u>zhrnutie</u>:
1. Mám D
2. Dopočítam K (Y-H veta)
3. $K = U \Lambda U^T$: spekt. rozklad $\Rightarrow$ $K = YY^T$ pre $Y \in \mathbb{R}^{n \times r}$
   $\approx YY^T$ pre $Y \in \mathbb{R}^{n \times q}$ $(q < r)$

• ak začnem s $x_1, \dots, x_n \in \mathbb{R}^m$ $\rightarrow$ dopočítam D $\Rightarrow$ PCA = KMŠ
↳ potom KMŠ pre D = PCA na pôvodných dátach

• dostanem akurát možno zrotované vektory súradníc $y_1, \dots, y_n$ voči prvým q hl. komponentom

- <u>rotácia</u>: máme $K = YY^T$, $Y \in \mathbb{R}^{n \times r}$

$\rightarrow$ vezmime $\tilde{Y} = Y \cdot Q$ $\Rightarrow$ $\tilde{Y}\tilde{Y}^T = YQQ^TY^T = YY^T = K$: Q ; Q ľub. rotácia

rotácia: máme $K = YY^T$, $Y \in \mathbb{R}$

$\rightarrow$ vezmime $\tilde{Y} = Y \cdot Q$ (ortog., $r \times r$) $\Rightarrow \tilde{Y}\tilde{Y}^T = Y\underbrace{QQ^T}_{I}Y^T = YY^T = K$: aj ľub. rotácia vyhovuje K

- ak Δ nie je presne maticou vzd. $\Rightarrow K$ nemusí byť k.s.d. : môže mať $\lambda_i < 0$
  - $\rightarrow$ záporné vl. č. určite neberieme
  - $\rightarrow$ ak prvých q vl. č. sú dostatočne veľké, získame dostatočne dobrú aproximáciu
    - časté pravidlo: také najmenšie q : $\dfrac{\lambda_1 + \ldots + \lambda_q}{|\lambda_1| + \ldots + |\lambda_n|} > 0{,}8$ (napr.)
  - $\rightarrow$ na zbavenie $\lambda_i < 0$ môžeme spraviť: ku všetkým D_{ij} $(i \neq j)$ pripočítame kladnú konštantu

2.) Metrické MŠ (Metric MS)

- zovšeobecnenie KMŠ

klasické: $D_{ij} \approx d_{ij}$, d_{ij} – vzd. medzi y_i a y_j

metrické: $f(D_{ij}) \approx d_{ij}$

f - monotónna funkcia, parametrizovaná

- napr. $f(D_{ij}) = \alpha + \beta D_{ij}$, $\alpha, \beta \geq 0$

- kvalita aproximácie pre dané $y_1, \ldots, y_n$ ($y_i \in \mathbb{R}^q$), parametre f : vážený súčet štvorcov

$$\sum_{i<j} w_{ij} \underbrace{(f(D_{ij}) - d_{ij})^2}_{\text{chyba fitu}} = S(Y)$$

w_{ij} – váhy (zvolené)

$S(Y)$ – stress function

úloha: $S(Y) \rightarrow \min_{y_1, \ldots, y_n, \text{ parametre } f}$

- nelineárna optimalizácia: numerické riešenie
- vhodný štartovací bod: výsledok KMŠ

- príklady váh: $w_{ij} = \dfrac{1}{d_{ij}^2}$, $w_{ij} = \dfrac{1}{\sum_{k<l} d_{kl}^2}$

- podobnosť s jadrovou PCA: dáta transformujeme a potom fitujeme

- podobnosť s jadrovou PCA: dáta transformujeme a potom fitujeme (kPCA je špec. prípad mMŠ za istých podmienok)

- Sammonovo mapovanie (Sammon mapping): $w_{ij} = \dfrac{1}{d_{ij} \sum_{k<l} d_{kl}}$, $f(D_{ij}) = D_{ij}$
  ↳ v 1969

$$\Rightarrow S(Y) = \frac{1}{\sum_{k<l} d_{kl}} \sum_{i<j} \frac{(D_{ij} - d_{ij})^2}{d_{ij}} \longrightarrow \min$$

→ dáva väčšiu váhu bližším objektom: snaží sa ich fitovať presnejšie

3.) Nemetrické škálovanie (Nonmetric MS)

- snažíme sa zachovať iba poradie vzdialeností (ranky)
  ↳ nezaujímajú nás presné vzdialenosti, alebo ich nepoznáme
- máme iba usporiadanie $s = \frac{n(n-1)}{2}$ vzdialeností: $D_{i_1 j_1} < \dots < D_{i_s j_s}$
- cieľ: ↑ toto aproximovať

1.) počiatočné $y_1, \dots, y_n \in \mathbb{R}^q$ – náhodné, pomocou kMŠ

- $d_{ij} = \| y_i - y_j \|$

→ chceme: $\underbrace{d_{i_1 j_1}}_{v_1} < \dots < \underbrace{d_{i_s j_s}}_{v_s}$

2.) d_{ij} aproximujem pomocou $\hat{d}_{ij}$

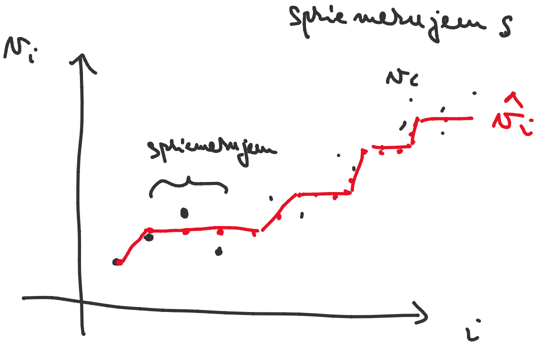
$v_i \approx \hat{v}_i$: $\hat{v}_1 \le \dots \le \hat{v}_s$

a) izotonická (monotónna) regresia (isotonic / monotonic regression)

- idem postupne po $v_1, v_2, \dots, v_s$
- ak $v_{i+1} \ge v_i$, tak sme spokojní
- ak $v_{i+1} < v_i \Rightarrow$ všetky $v_j, v_{j+1}, \dots, v_i$, ktoré sú väčšie ako v_{i+1}, spriemerujem s v_{i+1}.

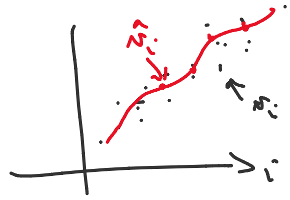
Príklad:

i	v_i	$\hat{v}_i$
1	1,2	1,2
2	1,4 ✓	~~1,4~~ 1,43
3	1,6 ✓	~~1,6~~ 1,43
4	1,3 ✗	1,43
⋮		



⇒ schodíkovitá fcia spĺňajúca $\hat{v}_1 \le \hat{v}_2 \le \dots \le \hat{v}_s$.

b) monotónne splajny : aproximujem $v_1, \dots, v_s$ splajn (kubický, kvadratický, ...) u ktorého vyžadujeme monotónnosť



3.) aktualizujem y_i, aby sa vylepšil stres: $S(Y) = \sum_{i<j} w_{ij}(d_{ij} - \hat{d}_{ij})^2 \to \min_{y_1, \dots, y_n}$

– nevypočítame rovno optimálne y_i – napr. iba 1 krok iteračného algoritmu

→ opakujem 2.) a 3.) až do konvergencie

<u>Voľba q</u> (pre mMŠ, nMŠ)

– 2 alebo 3 kvôli vykresleniu

– zakreslenie optim. hodnôt stresu pre rôzne q : S_q^*

→ hľadáme "lakeť" : hodnotu, po ktorej už pridávanie rozmerov nemá veľký prínos