RDD - 09

19 April 2021 14:47

• o týždeň: 26.4. – iba nahrávané prednáška
  - nie naživo

Minule: MŠ – máme vzdialenosti: $D = \begin{bmatrix} 0 & D_{12} & \cdots & D_{1n} \\ \vdots & & \ddots & \\ D_{n1} & \cdots & & 0 \end{bmatrix}_{n\times n}$

→ reprezentácia v q-rozmere

a) klasické MŠ

b) metrické MŠ – transformované vzdialenosti, váhy

c) nemetrické MŠ – iba poradie vzdialeností

Pr: americkí senátori: 15 senátorov – # rôznych hlasovaní

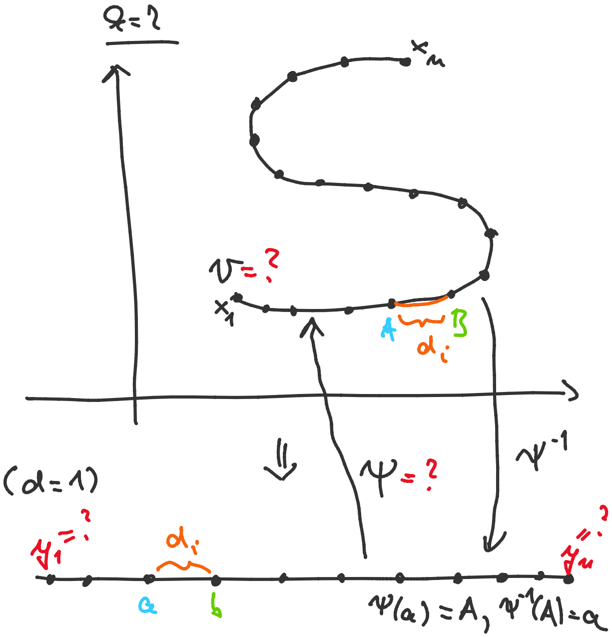
: priebeh korony – 1. rok, 31 krajín

→ korel. matica C ⇒ matica nepodobnosti: $D = 1 - C$

(všetky prvky vyšli nezáporné)

: $D = -\ln(C) = \log \frac{1}{C}$

Isomap



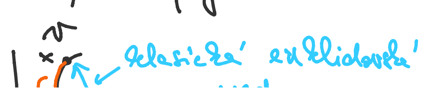
- isometric feature mapping
- menejrozmerná štruktúra, na ktorej ležia dáta
→ vnorenie (embedding) menšej štruktúry v $\mathbb{R}^q$

- menejrozmerná štruktúra: varieta (manifold)
voľne:
~ priestor, kt. sa lokálne podobá na euklidovský, ale môže byť zakrivený (krivka, kružnica, ...)

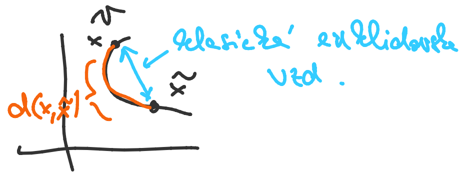
- $V \subseteq \mathbb{R}^q$, je d-rozmerná
  → môžeme ju zapísať ako $V = \Psi(M)$ pre nejaké $M \subseteq \mathbb{R}^d$, $\Psi: M \to V$
  → $y \in M \Leftrightarrow \Psi(y) \in V$
  → inverzné zobrazenie: Ψ^{-1}: $\forall y \in M\ \exists x \in V : y \in M^{-1}(x)$

• pekné vlastnosti: hladká – v každom bode má Ψ ľub. deriváciu
: konvexnosť – M je konvexná v $\mathbb{R}^d$

• metrika na variete: $d(x, \tilde{x})$ = dĺžka najkratšej krivky vo V, kt. spája x a $\tilde{x}$
pre $x, \tilde{x} \in V$



pre $x, \tilde{x} \in \mathcal{V}$



predp.: zobrazenie ψ je izometrické:

$$\forall y, \tilde{y} \in M : \underbrace{\|y - \tilde{y}\|}_{\text{klasická euk. vzd. v } \mathbb{R}^d} = \underbrace{d(\psi(y), \psi(\tilde{y}))}_{\text{vzd. na } \mathcal{V}}$$

Cieľ: máme $x_1, \dots, x_n \in \mathcal{V} \subseteq \mathbb{R}^q$, $\mathcal{V} = \psi(M)$, $M \subseteq \mathbb{R}^d$

nepoznáme $\mathcal{V}$ ani ψ

→ hľadáme $\mathcal{V}$, $\psi^{-1}(x_1), \dots, \psi^{-1}(x_n)$

$\psi^{-1}(x_1) = y_1, \dots, \psi^{-1}(x_n) = y_n \in \mathbb{R}^d$

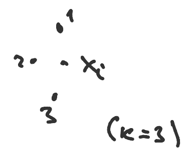
v praxi: $\hat{y}_1, \dots, \hat{y}_n \in \mathbb{R}^q$, zvyčajne $q \in \{2, 3\}$

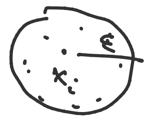
MŠ: zachovanie vzd. medzi objektami ⟷ Isomap: zachovanie $d(\ )$ – "varietných" vzdialeností

Algoritmus

1. Graf susednosti:

- vypočítame $D_{ij} = \|x_i - x_j\|$ $\forall i \neq j$
- $\forall x_i$ nájdeme jeho susedov
  - → a) pre zvolené ε: všetky body bližšie ako ε
  - → b) pre zvolené k: k najbližších bodov



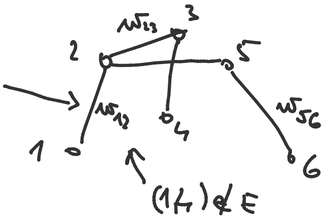


- vytvoríme vážený graf $G = (V, E, w)$ *,

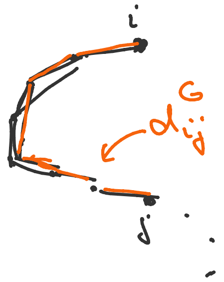
$V = \{1, \dots, n\}$, E: $(i,j) \in E$ ak i a j sú susedia (pre b) stačí ak aspoň jeden je susedom druhého)

$w_{ij} = D_{ij}$

váha na hrane (i,j)



V – vrcholy
E – hrany



2. geodetické vzdialenosti: $d_{ij} = d(x_i, x_j)$ – odhadneme

ako d_{ij}^G = najkratšia vzdialenosť medzi * i a j po grafe G

(váha w_{ij} = dĺžka hrany (i,j))

3. Mnohorozmerné škálovanie: $D^G = (d_{ij}^G)$ ...

3. Mnohorozmerné škálovanie: $D^G = (d_{ij}^G)_{i,j=1,\dots,n}$: $n \times n$ matica vzdialeností

→ na D^G aplikujem klas. škálovanie pre zvolené q ⇒ Výstup: $y_1, \dots, y_n \in \mathbb{R}^q$

↓

q-rozmerná reprezentácia dát, ktorá najlepšie zachytáva geometriu variety

Voľba q:

- $q \in \{2,3\}$ - vykreslenie
- lakťový graf (pomocou vl. čísel a kMŠ)
- $\dfrac{\lambda_1 + \dots + \lambda_q}{|\lambda_1| + \dots + |\lambda_n|} \geq$ konšt.

- matica vzdialeností $n \times n$ ⇒ problém s veľkým n

  → landmark isomap.

Pr.:

swiss roll - umelé dáta