RDD - 03

01 March 2021 13:43

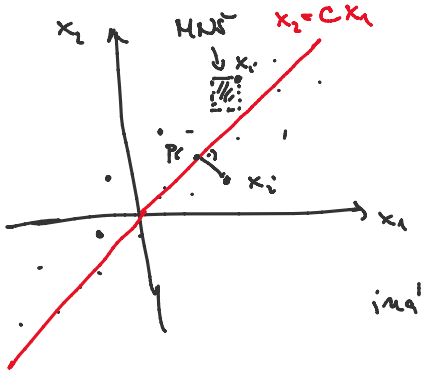
MNIST – ručne napísané čísla 0–9

.txt – 60 000 obrázkov : každý rozpixelovaný ako $28 \times 28 = 784$ hodnôt

⇓ $n = 60\,000$ – odtiene šedej ⇓ $q = 784$

– na každom obrázku je prečítané, aké číslo to je

PCA ako optimálne lineárne projekcie

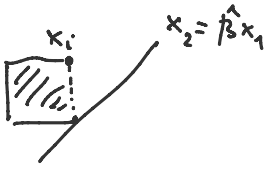


$x_1, \dots, x_n \in \mathbb{R}^q$ – centrované ($x_i - \bar{x}$)

$\boxed{q = 2}$ – regresná priamka?

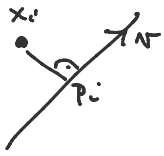
MNŠ: ↳ $x_2 = \hat{\beta} x_1$

$\hat{\beta}$: $\sum_{i=1}^{n} (x_{i,2} - \beta x_{i,1})^2 \to \min_{\beta}$



iná vzdialenosť : $\| x_i - p_i \|$

p_i – proj. x_i na priamku danú vektorom v



$$p_i = \frac{v v^T}{v^T v} x_i$$

$$\Rightarrow \sum_{i=1}^{n} \| x_i - \frac{v v^T}{v^T v} x_i \| \to \min_{v \in \mathbb{R}^2}$$

$\boxed{q > 2}$ p_i – proj. na nadrovinu = stĺpcový priestor matice $V \in \mathbb{R}^{q \times q}$

$$p_i = V (V^T V)^{-1} V^T x_i$$

$$\Rightarrow \sum_{i=1}^{n} \| x_i - V (V^T V)^{-1} V^T x_i \| \to \min_{\substack{V \in \mathbb{R}^{q \times (q-1)} \\ V \text{ reg.}}}$$

(nadrovinu chcem $(q-1)$-rozmernú)

dokonca:

$$\sum_{i=1}^{n} \| x_i - V V^T x_i \| \to \min_{\substack{V \in \mathbb{R}^{q \times q} \\ V \text{ ortog. stĺpce}}}$$

MNŠ:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Riešenie: V : prvých q vl. vektorov matice $X^T X \approx S$

$$S = \frac{1}{n-1} \sum (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{n-1} \sum_{i=1}^{n} x_i x_i^T = \frac{1}{n-1} \underbrace{(x_1, \dots, x_n)}_{X^T} \underbrace{\begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}}_{X} = \frac{1}{n-1} X^T X$$

(↑ cent.)

$\Rightarrow$ V určuje priečky q hlavných komponentov

Projekčné sledovanie – projection pursuit (PP)

- hľadanie štruktúry v dátach (grafické)
  → zaujímavé 1/2/3-rozmerné projekcie
- Kruskal (1969), Friedman & Tukey (1972)
  ↳ meno PP: „A projection pursuit algorithm for exploratory data analysis"
  ↳ často je to pripisované im

- zvolíme index φ, ktorý vyjadruje mieru zaujímavosti projekcie
  : každú proj. ohodnotíme číslom φ
  → maximalizujeme tento index

⇒ PP: 1. zvoliť φ
2. nájsť maximá (lok. a glob.) cez všetky q-rozmerné projekcie ($q = 1/2/3$)

Projekčné indexy:
- je vhodné ak sú afinne invariantné: vzhľadom na natáčanie a škálovanie
  → nieši sa to: štandardizované dáta (rescaled): priemer 0_q
  variancia I_q
  ↳ budeme to predpokladať { populácia: $E(X) = 0$, $Var(X) = I$
  { dátovo: $\bar{x} = 0$, $S = I$

- dosiahneme to: 1.) centrovanie: $\tilde{x}_i = x_i - \bar{x}$
  2.) $S = \frac{1}{n(-1)} \tilde{X}^T \tilde{X} = U \Lambda U^T$: spekt. rozklad
  $\Rightarrow S^{-1} = U \Lambda^{-1} U^T = \underbrace{U \Lambda^{-1/2}}_{V} \underbrace{\Lambda^{-1/2} U^T}_{V^T}$

$\tilde{\tilde{x}}_i = V^T \tilde{x}_i$, $i = 1, \dots, n$ } maticovo: $\begin{pmatrix} \tilde{\tilde{x}}_1^T \\ \vdots \\ \tilde{\tilde{x}}_n^T \end{pmatrix} = \begin{pmatrix} (V^T \tilde{x}_1)^T \\ \vdots \\ (V^T \tilde{x}_n)^T \end{pmatrix} =$

$\tilde{\tilde{X}} = \begin{pmatrix} \tilde{x}_1^T V \\ \vdots \\ \tilde{x}_n^T V \end{pmatrix} = \tilde{X} V = \tilde{X} U \Lambda^{-1/2}$

teoreticky: $\Sigma = Var(X) = U \Lambda U^T$ $\quad U \Lambda^{-1/2}$
$\Rightarrow Var(V^T X) = V^T Var(X) V =$
$= \Lambda^{-1/2} U^T \cdot U \Lambda U^T \cdot U \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda^{1/2} \Lambda^{1/2} \Lambda^{-1/2} = I$

→ tedy počítame s $\tilde{\tilde{X}} = \tilde{X} U \Lambda^{-1/2}$ (tie budeme značiť X)

PCA: $Y = \tilde{X} U$ ⇒ štandardizovanie je fakticky PCA, akurát ešte prekĺbujeme pomocou $\Lambda^{-1/2}$ variancie na 1.

---

normálne dáta sú "nudné"
→ väčšina projekcií mnohorozmerných dát vyzerá približne normálne
→ normálne rozdelenia majú (všetky) marginálne znova normálne
⇒ indexy často vyjadrujú mieru nenormality

→ chceme $\varphi(Y) = 0$ pre $Y \sim N(\,)$
$\varphi(Y) \geq 0 \quad \forall Y$
ideálne: $\varphi(Y) > 0$ pre $Y \not\sim N$

1-rozmer: $(q = 1)$: teoretický: projekcia: $Y = h^T X$, $h \in \mathbb{R}^q$
(Y: 1-rozmer, X: q-rozmer)

$$Y = h_1 X_1 + \dots + h_q X_q$$

výberovo: $y_i = h^T x_i$, $i = 1, \dots, n$

$$(y_1, \dots, y_n) = (h^T x_1, \dots, h^T x_n) = h^T (x_1, \dots, x_n)$$
$$y^T \qquad y^T = h^T X^T \qquad X^T$$
$$y = X h$$

φ: 1.) variancia: $\varphi(Y) = D(Y) = D(h^T X) \to \max_h$: to maximalizuje PC1

⇒ pre takéto φ je PP = PCA

2.) momenty

a) špicatosť: $\varphi(Y) = |\mathrm{kurt}(Y) - 3| \to \max$

$$\mathrm{kurt}(Y) = E\left[\left(\frac{Y - E(Y)}{\sqrt{D(Y)}}\right)^4\right] : \text{špicatosť}$$

$$\mathrm{kurt}(N(0,1)) = 3$$

- štandardizované dáta: $\mu_1 = E(Y) = 0$
$\mu_2 = E(Y^2) = 1$

- pracuje sa s kumulantmi (nie s momentami) – niečo ako momenty

- pracuje sa s kumulantmi (nie s momentami) - niečo ako momenty

↳ lebo kumulanty sú všetky nulové pre $N(0,1)$

: $\varkappa_j$ - j-ty kumulant Pre $\mu_1 = 0, \mu_2 = 1$: $\varkappa_3 = \mu_3$

: μ_j - j-ty moment : $E(Y^j)$ $\varkappa_4 = \mu_4 - 3$

$\Rightarrow$ maximalizujeme $\varphi(Y) = |\varkappa_4| = |\mu_4 - 3|$

b) zložitejšie kombinácie : napr. $\varphi(Y) = \frac{\varkappa_3(Y)^2}{12} + \frac{\varkappa_4(Y)^2}{48} \rightarrow \max$

výberovo : odhady : $\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} y_i^j$, alebo iný odhad momentov

$\Rightarrow \hat{\varkappa}_3 = \hat{\mu}_3$, $\hat{\varkappa}_4 = \hat{\mu}_4 - 3$

- výpočtovo nenáročný, tradičný postup
- veľmi nerobustný : outliery prudko ovplyvňujú $\hat{\varkappa}_4$

maximalizácia φ - nekonkávna f., zvyčajne veľa lok. maxím $\Rightarrow$ zložité

- treba použiť algoritmy, kt. sa nezaseknú v lok optime
- multistart

napr.:
- genetický alg.,
particle swarm optimization,
...
→ Stoch. optim. metódy (Harman)

praktické použitie :

→ každé lok. max. môže predstavovať zaujímavú projekciu ↑ výhoda
→ spustíme alg. veľakrát, zapamätáme si výsledky
→ vyhľadáme čo najrozdielnejšie proj. z tých získaných { výrazné poklesy $\varphi(Y)$ / nuly