§ 2 Dünne Hauptkomponentenanalyse

(sparse PCA)

$X \in \mathbb{R}^{D \times N}$ Datenmatrix, M Zieldimension (klein) vorgegeben.

<u>PCA</u>: Berechne Hauptkomponenten $b^1, \dots, b^M \in \mathbb{R}^D$.

<u>Problem</u>: Oft ist D groß, möchte aber Daten X erklären aber mit nur wenigen (k) Einträgen in den b^i's.

<u>Idee</u>: Ersetze b^1 durch optimale Lösung des Optimierungsproblems (und berechne danach sukzessive $b^2, \dots, b^M$).

$$(*) \qquad \begin{aligned} &\max \ z^T \Sigma z \\ &\|z\| = 1 \\ &\boxed{\|z\|_0 \leq k} \end{aligned},$$

wobei $\Sigma = X X^T$,

$$\|z\|_0 = |\{ i \in \{1, \dots, D\} : z_i \neq 0 \}|.$$

Nächstes Problem: (*) ist NP-schwer.

(→ Natarajan, 1995)

Ausweg: SDP-Relaxarierung (nach d'Aspremont et al. 2007)

Satz

$$(*) = \max_{\substack{Z \in S^D_{\geq 0} \\ \operatorname{Tr}(Z) = 1 \\ \|Z\|_0 \leq k^2 \\ \operatorname{rang} Z = 1}} \langle \Sigma, Z \rangle \quad (**) \qquad \leq \qquad \max_{\substack{Z \in S^D_{\geq 0} \\ \operatorname{Tr}(Z) = 1 \\ \langle J, |Z| \rangle \leq k}} \langle \Sigma, Z \rangle \quad (***),$$

dabei $J = \begin{bmatrix} 1 & \dots & 1 \\ & \ddots & \\ 1 & \dots & 1 \end{bmatrix}$ und das rechte Maximierungsproblem heißt SDP-Relaxierung der sparse PCA.

Das kann effizient (in Theorie & Praxis) gelöst werden.

(siehe Link "Literatur / Software" der VL Convex Optimization).

Bew.: zum ersten "=": Sei z optimal für (*), dann ist $Z = zz^T$ optimal für (**), und umgekehrt.

Zum zweiten "$\leq$":

Sei $Z = zz^T$ optimal für $(**)$, dann ist $\|z\|_0 \leq k$,

OBdA. $z_{k+1} = \dots = z_D = 0$. Dann ist

$$\langle J, |Z| \rangle = \sum_{i,j=1}^{D} |Z_{ij}|$$

$$= \sum_{i,j=1}^{D} |z_i z_j|$$

$$= \sum_{i,j=1}^{D} |z_i|\,|z_j|$$

$$= \left(\sum_{i=1}^{k} |z_i| \right)^2$$

$$\overset{\text{Cauchy-Schwarz}}{\leq} \|z\| \cdot \left\| \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\|, \qquad \|z\| = 1,$$

$$= k \qquad \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^k.$$

$\boxtimes$

Verwendung von $(***)$ zur Bestimmung einer sparse PCA:

1) Bestimme optimale Z
2) Bestimme Eigenvektor z zu $\lambda_{max}(Z)$.
3) $\Sigma \leftarrow \Sigma - z^T \Sigma z \, zz^T$.
4) Gehe zu 1). bis $\Sigma \approx 0$.

Zur Modellierung von $\langle J, |Z| \rangle \leq k$:

$$\langle J, |Z| \rangle = \underbrace{\sum_{i=1}^{D} \sum_{j=1}^{D} |Z_{ij}|}_{\ell_1\text{-Norm}} \leq k .$$

Für $x \in \mathbb{R}^n$ gilt

$$\|x\|_1 \leq k \iff x \in \operatorname{conv}\{\pm k e_i : i = 1, \dots, n\}$$

$$\iff \exists \lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_n \geq 0 :$$

$$\sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} \mu_i = 1$$

$$x = \sum_{i=1}^{n} \lambda_i k e_i + \sum_{i=1}^{n} \mu_i k(-e_i)$$

D.h. man kann die Bedingung $\|x\|_1 \leq k$ mit linear vielen linearen Ungleichungen modellieren.

Naiv bräuchte man exponentiell viele:

$$\sum_{i=1}^{n} \pm x_i \leq k .$$