PCA und SVD

SVD $\hat{=}$ singular value decomposition (Singulärwert-zerlegung)

<u>Satz</u> Sei $A \in \mathbb{R}^{m \times n}$. Es gibt $U \in O(m)$, $V \in O(n)$ orthogonale Matrizen mit

$$U^T A V = \operatorname{diag}(\sigma_1, \dots, \sigma_p),$$

wobei $p = \min\{m, n\}$ und $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

<u>Bew.</u>: ($\rightarrow$ Numerik)

Wähle $v_1 \in \mathbb{R}^n$ mit $\|v_1\| = 1$ und

$$\|Av_1\| = \max_{\|x\|=1} \|Ax\| = \|A\| = \sigma \geq 0$$

Sei $u_1 \in \mathbb{R}^m$ mit $\|u_1\| = 1$ und

$$Av_1 = \sigma u_1$$

Ergänze u_1 zu einer ONB von $\mathbb{R}^n$: $U = [u_1\, u_2 \dots u_n] \in O(n)$
und v_1 zu einer ONB von $\mathbb{R}^m$: $V = [v_1\, v_2 \dots v_m] \in O(m)$.

Dann ist $\underbrace{U^T A V}_{A'} = \begin{bmatrix} \sigma & w^T \\ 0 & B \end{bmatrix}$ mit $w \in \mathbb{R}^{n-1}$, $B \in \mathbb{R}^{(m-1)\times(n-1)}$.

Beh.: $w = 0$.

Bew.: Es gilt $\|A\| = \|A'\| = \sigma$.

Außerdem $\left\|A'\begin{pmatrix}\sigma\\ w\end{pmatrix}\right\|^2 = \left\|\begin{pmatrix}\sigma^2 + w^T w\\ *\end{pmatrix}\right\|^2$

$$\geq (\sigma^2 + w^T w)^2,$$

und $\left\|\begin{pmatrix}\sigma\\ w\end{pmatrix}\right\|^2 = \sigma^2 + w^T w$.

Also $\underbrace{\|A'\|^2}_{= \sigma^2} \geq \sigma^2 + w^T w$, und somit $w^T w = 0$.

Also $U^T A V = \begin{bmatrix}\sigma & 0\\ 0 & B\end{bmatrix}$ und man kann induktiv mit $B \in \mathbb{R}^{(m-1)\times(n-1)}$ weitermachen. $\boxtimes$

Sei $X \in \mathbb{R}^{D\times N}$. Bestimme eine SVD von X:

$$U^T X V = \operatorname{diag}(\sigma_1, \dots, \sigma_p), \qquad U \in O(D),\ V \in O(N).$$

Also $X = U \operatorname{diag}(\sigma_1, \dots, \sigma_p) V^T$ und

$$X X^T = U \operatorname{diag}(\sigma_1, \dots, \sigma_p) V^T V \operatorname{diag}(\sigma_1, \dots, \sigma_p) U^T$$

$$= U \begin{bmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2\end{bmatrix} U^T.$$

D.h. die Zeilenvektoren der Matrix U enthalten die Hauptkomponenten von X.

Die $\sigma_1, \ldots, \sigma_p$ heißen die <u>Singulärwerte</u> von X:

$$\sigma_{max}(X) = \sigma_1(X) = \sigma_1, \ldots, \quad \sigma_{min}(X) = \sigma_p(X) = \sigma_p.$$

Ang. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$.

Seien u_i die Spaltenvektoren von U, v_i die von V. Dann ist

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

eine <u>Singulärwertzerlegung</u> von A.

Gute numerische Algorithmen zur Berechnung der SVD:

$\longrightarrow$ Buch Golub, van Loan – Matrix Computations.

SVD und Rang-k-Approximation

<u>Satz</u> Sei $A \in \mathbb{R}^{m \times n}$ und $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ eine Singulärwertzerlegung. Sei $0 \leq k < r$. Dann

$$\min_{B \in \mathbb{R}^{m \times n},\ \operatorname{rang} B \leq k} \|A - B\| = \|A - \underbrace{\sum_{i=1}^{k} \sigma_i u_i v_i^T}_{A_k}\| = \sigma_{k+1}$$

Bew.: klar: • $\operatorname{rang} A_k = k$

• $\| A - A_k \| = \sigma_{k+1}$.

Beh.: Sei $B \in \mathbb{R}^{m\times n}$ mit $\operatorname{rang} B \le k$. Dann gilt $\| A - B \|^2 \ge \sigma_{k+1}^2$.

Bew.: Wir haben $\dim \ker B \ge n-k$. Also $\operatorname{span}\{v_1, \dots, v_{k+1}\} \cap \ker B \ne \{0\}$. Sei $z \in \mathbb{R}^n$ in diesem Durchschnitt mit $\|z\| = 1$. Dann gilt

$$\|Az\|^2 = \sum_{i=1}^{r} \sigma_i^2 (v_i^T z)^2 \ge \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2$$
$$\ge \sigma_{k+1}^2 \sum_{i=1}^{k+1} (v_i^T z)^2 = \sigma_{k+1}^2 .$$

Also

$$\|A - B\|^2 \ge \|(A-B)z\|^2 = \|Az\|^2 \ge \sigma_{k+1}^2 . \qquad \boxtimes$$

SVD und Geometrie der Konditionszahl

Def.: Sei $A \in \mathbb{R}^{n\times n}$, $\operatorname{rang} A = n$. Die Konditionszahl von A ist

$$\kappa(A) = \|A\| \cdot \|A^{-1}\| = \sigma_{max}(A) / \sigma_{min}(A).$$

Def.: $\Sigma = \{ A \in \mathbb{R}^{n\times n} : \text{rang } A < n \}$

ist die Menge der singulären Matrizen.

$$d_F(A, \Sigma) = \min \{ \|A - B\|_F : B \in \Sigma \}$$

(Frobenius-) Abstand der Matrix A zu singulären Matrizen.

$\left(\|Z\|_F = \sqrt{Tr(Z^T Z)} \right)$.

Satz Sei $A \in \mathbb{R}^{n\times n}$. Dann gilt [Das Theorem von Eckart-Young]

$$d_F(A, \Sigma) = \sigma_{min}(A).$$

Insbesondere gilt (falls rang $A = n$)

$$\kappa(A) = \frac{\|A\|}{d_F(A, \Sigma)}.$$

(D.h. die Konditionszahl ist <u>invers proportional</u> zum Abstand der singulären Matrizen)

Bew.: Definiere $d(A, \Sigma) = \min \{ \|A - B\| : B \in \Sigma \}$.

Nach Rang-(n-1)-Approximationsatz ist klar, dass $d(A,\Sigma) = \sigma_{min}(A)$ gilt.

Allgemein gilt $\|A\|_F \geq \|A\|$, also $d_F(A,\Sigma) \geq d(A,\Sigma)$. Bleibt $d_F(A,\Sigma) \leq d(A,\Sigma) = \sigma_{min}(A)$ zu zeigen: Sei $A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$ SVD von A. Definiere $B = \sum_{i=1}^{n-1} \sigma_i u_i v_i^T \in \Sigma$. Dann $A - B = \sigma_n u_n v_n^T$ und

$$\|A-B\|_F = \sqrt{Tr\left((\sigma_n u_n v_n^T)^T(\sigma_n u_n v_n^T)\right)}$$

$$= \sqrt{Tr\left(\sigma_n^2 v_n u_n^T u_n v_n^T\right)}$$

$$= \sigma_n = \sigma_{min}(A).$$

$\boxtimes$