

Recommendation Systems II

Ratingmatrix $A = (a_{ij})$

Nutzer n  Movies m

Bewertung des j -ten Movie durch den i -ten Nutzer

Features $r = \text{Rang}(A) = \underline{\text{Rang der Matrix A}}$
max. Anzahl linear unabhängiger

Spalten bzw. Zeilen von A

$$A \cdot x = \underbrace{\begin{pmatrix} a^{(1)} & a^{(2)} & \dots & a^{(m)} \end{pmatrix}}_{\text{Spalten von A}} \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}}_{\text{Linearkombination von Spalten}} = x_1 \cdot a^{(1)} + x_2 \cdot a^{(2)} + \dots + x_m \cdot a^{(m)}$$

jede LK aus r l.u. darstellbar =

$$\dim \text{Bild}(A) = r$$

Dimension

Bild(A)

Bewertung von Features durch Nutzer

Gewichtung von Features

Ausprägung von Features in Movies

Singularwertzerlegung (SVD): $A = U \cdot \Sigma \cdot V^T$

$n \times m$ $n \times r$ $r \times r$ $r \times m$

wobei $U = (u_1, \dots, u_r)$ mit $u_i^T \cdot u_j = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

orthonormale Matrizen

$$V = \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix}$$

mit $v_i^T \cdot v_j = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

Kurz: $U^T \cdot U = \begin{pmatrix} u_1^T \\ \vdots \\ u_r^T \end{pmatrix} \cdot (u_1, \dots, u_r) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = I_r$ (Identität)

$$V \cdot V^T = \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix} \cdot (v_1, \dots, v_r) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = I_r$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r) \quad \text{mit} \quad \sigma_i > 0, i=1, \dots, r$$

Singularwerte

Approximation von A

(2)

• Frobenius-Norm: $\|A\|_F := \sqrt{\sum_{j=1}^m \sum_{i=1}^n a_{ij}^2} = \sqrt{\text{tr}(A^T A)} = \sqrt{\text{tr}(A \cdot A^T)}$

$\sum_{j=1}^m b_{jj} = (A^T A)_{jj} = \sum_{i=1}^n a_{ij} \dots a_{ij} = \sum_{i=1}^n a_{ij}^2$ (j-te Spalte)

$\text{tr}(B) = \sum_{j=1}^m b_{jj}$ (Summe auf der Hauptdiag.)

$\|A\|_F^2 = \text{tr}(A^T A) \stackrel{\text{SVD}}{=} \text{tr}(V^T \Sigma^T U^T U \Sigma V) = \text{tr}(\Sigma^2) = \sigma_1^2 + \dots + \sigma_r^2$

$A = U \cdot \Sigma \cdot V$

Summe der Singulärwerte
 $\sigma_i^2 = d_i, i=1, \dots, r$: "d1 + ... + dr"
 Summe der Eigenwerte: Spur von B

• Suche nach der besten Approximation von A vom Rang k

abgeschätzte Anzahl von Features

$\min_{\text{Rang}(A_k)=k} \|A - A_k\|_F^2$

o.B.d.A. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ (aufsteigend sortiert)

$A_k = U \cdot \Sigma_k \cdot V$, wobei $\Sigma_k = \begin{pmatrix} \sigma_1 & \dots & \sigma_k & 0 & \dots & 0 \\ & \ddots & & & & \\ & & \sigma_{k+1} & \dots & \sigma_r & \\ & & & \ddots & & \\ & & & & 0 & \dots & 0 \end{pmatrix}$ (k größte Singulärwerte)

$\Rightarrow \|A - A_k\|_F^2 = \|U \cdot \Sigma \cdot V - U \cdot \Sigma_k \cdot V\|_F^2 = \|U \cdot (\Sigma - \Sigma_k) \cdot V\|_F^2 = \text{tr}((\Sigma - \Sigma_k)^2) = \text{tr} \left(\begin{pmatrix} \sigma_1 - \sigma_k & & & \\ & \ddots & & \\ & & \sigma_{k+1} & \\ & & & \ddots & \\ & & & & \sigma_r \end{pmatrix} \right) = \sigma_{k+1}^2 + \dots + \sigma_r^2$

"Hauptkomponentenanalyse"

r-k kleinste Singulärwerte

"Nur am größten gewichtete Features betrachten!"

(falls $k=r \Rightarrow A_k = A$ und alle Features erhalten geblieben)

$A_k = U_k \cdot \Sigma_k \cdot V_k$ - SVD für $A_k \Rightarrow \text{Rang}(A_k) = k$.
 (nur erste k Spalten/Zeilen von U bzw. V)

Matrizenfaktorisierung

A - evtl. dünnbesetzt, d.h. manche Einträge fehlen

$$\min_{\text{Rang}(A_k)=k} \|A - A_k\|_F^2 = \sum_{i,j} (a_{ij} - a_{ij}^k)^2$$

Summe nur über vorhandene Bewertungen

Ansatz: $A_k = L \cdot R$ - Faktorisierung "Links-Rechts"

generisch: voller $n \times k$ mit $k < n, m$

↙ Rang ↘

$$\text{Rang}(L) = \text{Rang}(R) = k \Rightarrow \text{Rang}(A_k) = k \quad \text{Zulässigkeit}$$

Denn: Bild(R) - k-dim. $\Rightarrow z_i = R \cdot y_i, i=1, \dots, k$ l.u.

$$\text{Ang. } x_i = L \cdot z_i, i=1, \dots, k \text{ l.a.} \Rightarrow 0 = \sum_{i=1}^k d_i \cdot x_i = L \left(\sum_{i=1}^k d_i \cdot z_i \right)$$

$$= L \cdot R \cdot y_i \in \text{Bild}(L \cdot R) \quad \exists d \neq 0 \quad n \times k \mid 1$$

$$\Rightarrow \sum_{i=1}^k d_i \cdot z_i = 0 \Rightarrow d_1 = \dots = d_k = 0 \quad \text{Ver.} \quad \nexists d \neq 0$$

$$\text{Also: } \dim \text{Bild}(L \cdot R) = k \Rightarrow \text{Rang}(L \cdot R) = k$$

$\underbrace{\quad}_{=A_k}$

$$\min_{L, R} \sum_{i,j} \left(a_{ij} - \sum_{s=1}^k l_{is} \cdot r_{sj} \right)^2$$

unrestringiertes Optimierungsproblem

$f(l_{is}, r_{sj})$ $n \cdot k + k \cdot m = k(n+m)$ Variablen

(vgl. $A: n \times m$ n.m Variablen, i.a. $n \cdot m > k(n+m)$)

BIG DATA

z.B. $A = \begin{pmatrix} 5 & 3 & -1 \\ 4 & - & -1 \\ 1 & 1 & -5 \\ 1 & - & 4 \\ - & 1 & 5 & 4 \end{pmatrix}$

k=2

Gradientenverfahren bzgl. L und R ... (nach einigen Iterationen)

Prognose ↗

↙ Rekonstruktion

ähnlich $L \cdot R = A_k \approx$ ähnlich

N1	4.99	3.0	3.39	1.01
N2	4.0	3.18	2.98	1.01
N3	1.02	0.96	5.54	4.97
N4	1.0	0.6	4.78	3.97
N5	1.53	1.05	4.94	4.03
	M1	M2	M3	M4

$\Rightarrow$

N2	M2
N4	M3

nicht triviale Empfehlung

SV-D für $A_k = U_k \cdot \Sigma_k \cdot V_k$

2 große Singulärwerte

13.686, 5.98, 0.55, 0.15

kleine Singulärwerte