Feb 11, 2020

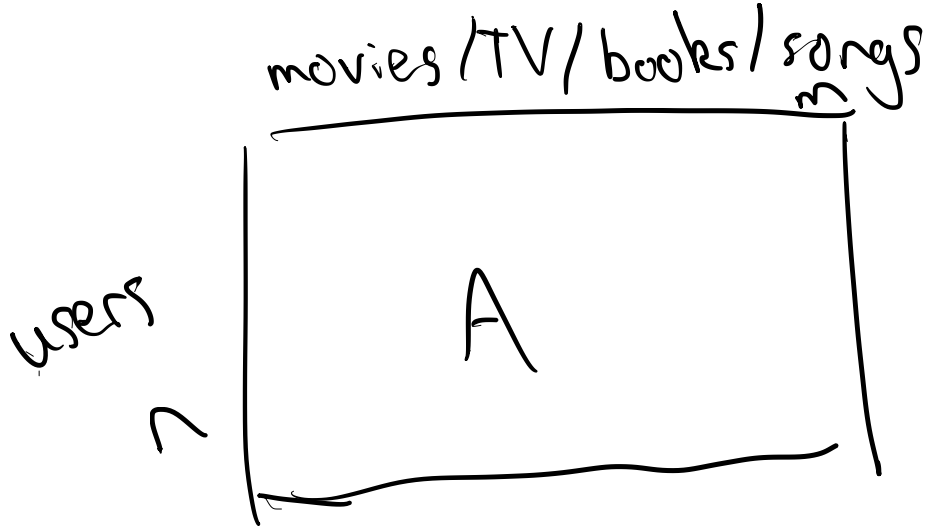
Last time: PCA, robust PCA

truncated SVD

Today: latent factor models thru matrix completion

HW1 due Thurs 11:59pm ET CMS

Recommender systems



$A_{ij} \approx$ how much user i likes song j (# plays)

song $j \in \mathbb{R}^k \Rightarrow y_j$

user $i \in \mathbb{R}^k \Rightarrow x_i$

$(y_j)_1 \approx$ how much "rap"

$(y_j)_2 \approx$ length

$A_{ij} \approx x_i^T y_j$

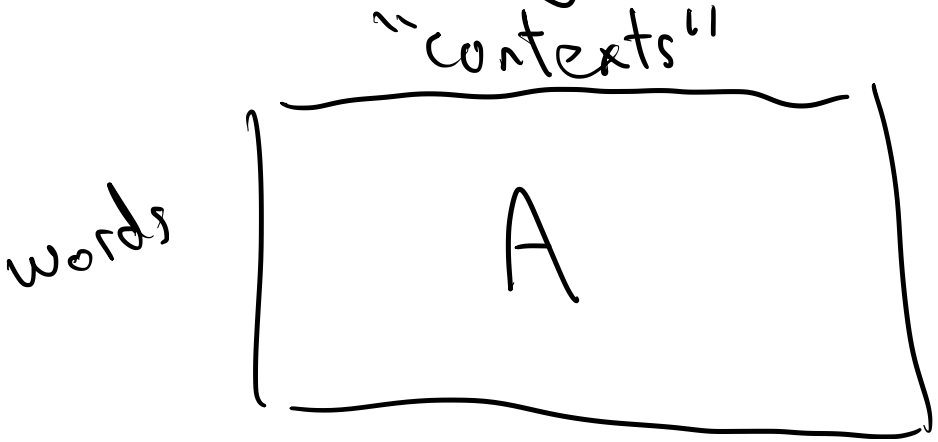
$(x_i)_1 \approx$ proclivity for rap

$(x_i)_2 \approx$ short songs

should user r like song s? $x_r^T y_s$

$A \approx XY^T \quad X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \quad Y^T = [y_1 \cdots y_m]$

Word embeddings



context $j \in \mathbb{R}^k \Rightarrow y_j$

word $i \in \mathbb{R}^k \Rightarrow x_i$

want: $\Pr(\text{choose } i \mid \text{context } j) \approx \dfrac{\exp(x_i^T y_j)}{\sum_r \exp(x_r^T y_j)}$

$$A_{ij} = \log(\text{Prob}(\text{choose } i \mid \text{context } j))$$

$$= x_i^T y_j - \log\left(\sum_r \exp(x_r^T y_j)\right)$$

constant per column
force to zero

$$A \approx XY^T \qquad X = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} \qquad Y^T = [y_1 \cdots y_M]$$

$$A \approx XY^T \qquad \min_{X,Y} \|A - XY^T\|_F^2$$

$$X \in \mathbb{R}^{n \times k} \quad Y \in \mathbb{R}^{m \times k}$$

Idea: truncated SVD

One problem: X, Y^T not unique

$$X = U_k \sqrt{\Sigma_k} \qquad Y = \sqrt{\Sigma_k} V_k^T$$

$$(XB)(B^{-1}Y^T) = XY^T$$

Add some regularization:

$$\min_{X,Y} \frac{1}{2}\|A - XY^T\|_F^2 + \lambda\left(\|X\|_F^2 + \|Y\|_F^2\right)$$

Major problem: only observe a few entries of A

$$\min_{X,Y} \frac{1}{2}\|P_\Omega(A - XY^T)\|_F^2 + \lambda(\|X\|_F^2 + \|Y\|_F^2)$$

$$\|P_\Omega(M)\|_F^2 = \sum_{(i,j)\in\Omega} M_{ij}^2$$

Alternating scheme

Fix Y_1 opt X

$$X^{k+1} = \min_X \frac{1}{2}\|P_\Omega(A - XY^T)\|_F^2 + \lambda\|X\|_F^2$$

ith row of A

$J_i = \{j \mid (i,j)\in\Omega\}$

$$\frac{1}{2}\sum_{j\in J_i}(A_{ij} - x_i^T y_j)^2$$

$$\|Y_{J_i} x_i - A_{i,J_i}\|_2^2 + \lambda\|x_i\|_2^2$$

l_2-reg LLS on each row

$$\left\| \begin{pmatrix} Y_{J_1} Y_{J_2} \cdots Y_{J_n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} A_{1J_1} \\ \vdots \\ A_{nJ_n} \end{pmatrix} \right\|_2^2 + \lambda \left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_2^2$$

Fix X_1 opt Y

$$Y^{k+1} = \min_Y \frac{1}{2} \| P_\Omega (A - X^{k+1} Y^T) \|_F^2 + \lambda \|Y\|_F^2$$

$\Omega^T = \{(j,i) \mid (i,j) \in \Omega\}$

$$\frac{1}{2} \| P_{\Omega^T} (A^T - Y X^T) \|_F^2 + \lambda \|Y\|_F^2$$

same!

Nuclear norm approach

$\sum_i \sigma_i$

$$\min_Z \frac{1}{2}\|P_\Omega(A-Z)\|_F^2 + \lambda\|Z\|_*$$

Last lecture: $\arg\min_Z \frac{1}{2}\|A-Z\|_F^2 + \lambda\|Z\|_*$

$$S_\lambda(A) = U S_\lambda(\Sigma) V^T = \sum_{i=1} u_i v_i^T \max(\sigma_i - \lambda, 0)$$

$$\simeq U_k \Sigma_k V_k^T$$

Idea: "fill in" A with Z,

use prox,

iterate

$$\bar{A}_{ij} = \begin{cases} A_{ij} & (i,j) \in \Omega \\ Z^k_{ij} & \text{o/w} \end{cases}$$

$$\bar{A} = Z^k + Q_\Omega(A - Z^k)$$

puts zero if $(i,j) \notin \Omega$

$$Z^{k+1} = \min_Z \frac{1}{2}\|\bar{A} - Z\|_F^2 + \lambda\|Z\|_* \quad (\star)$$

$$S_\lambda(\bar{A}) = S_\lambda\left(Z^k + Q_\Omega(A - Z^k)\right)$$

SGD

$$\min_{X,Y} \frac{1}{2}\|P_\Omega(A - XY^T)\|_F^2 + \lambda\left(\|X\|_F^2 + \|Y\|_F^2\right)$$

$$\sum_{(i,j)\in\Omega} \underbrace{(a_{ij} - x_i^T y_j)^2 + \frac{\lambda}{|J_i|}\|x_i\|_2^2 + \frac{\lambda}{|I_j|}\|y_j\|_2^2}_{f_{ij}(x,y)}$$

$$J_i = \{(i,j)\in\Omega\} \quad I_j = \{(i,j)\in\Omega\}$$

$$\min \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} f_{ij}(x,y)$$

$\nabla_x f_{ij}$ non-zeros on x_i

$\nabla_y f_{ij}$ " " y_i

sparse gradients

$O(k)$ parameters to update

Before: $A_{ij} \approx x_i^T y_j$ $\quad A \approx XY^T$

Maybe: $A_{ij} \approx x_i^T y_j + b_i + c_j + \mu$

overall bias

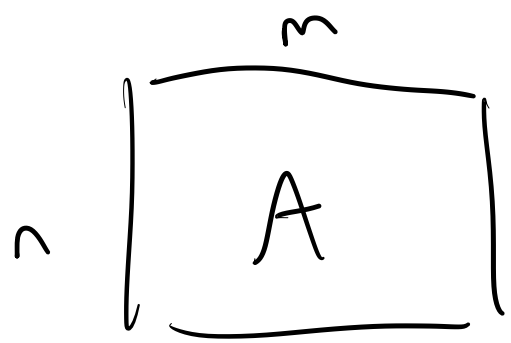
user i bias

item j bias

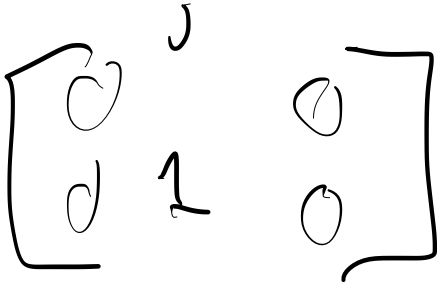
$$A \approx XY^T + 1b^T + c1^T + \mu 11^T$$

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (b_1 \cdots b_n)$$

Should this actually work?



$m \geq n$

$A = e_i e_j^T =$ 

$A = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ u_i, v_i sampled uniformly

$O(m^{5/4} \log(m) k)$ UAR samples

Thm: $\min \|Z\|_*$

s.t. $Z_\Omega = A_\Omega$

recover A w.h.p. (Candès & Recht)

(in)coherence

$$\max_i \|(UU^T)(i,:)\|_2$$

$$\|UU^T\|_{2\to\infty}$$