Topics:

1. Recap: Content-based vs collaborative filtering
2. Matrix Factorization : PCA & SVD
3. Generic MF for matrix-completion
4. NMF & optimization
5. clustering as MF
6. MF for feature-engineering → word-vectors; Eigen Faces
7. Hyper-param tuning
8. Netflix-Prize solution

optimization ✓

Recap

user-vec or item-vec

Content-based

- ✓ → User-user sim
- ✓ → Item-item sim
- → cold-start problem

user or item

Collaborative filtering

$A_{n \times m}$: ratings

users

items

Matrix Factorization (MF)

(decomposition)

Math

$12 = \underline{2} \times \underline{6}$

$12 = \underbrace{2 \times 3 \times 2}$

factors

algebra: $\underline{6} = \underline{2} \times \underline{3}$

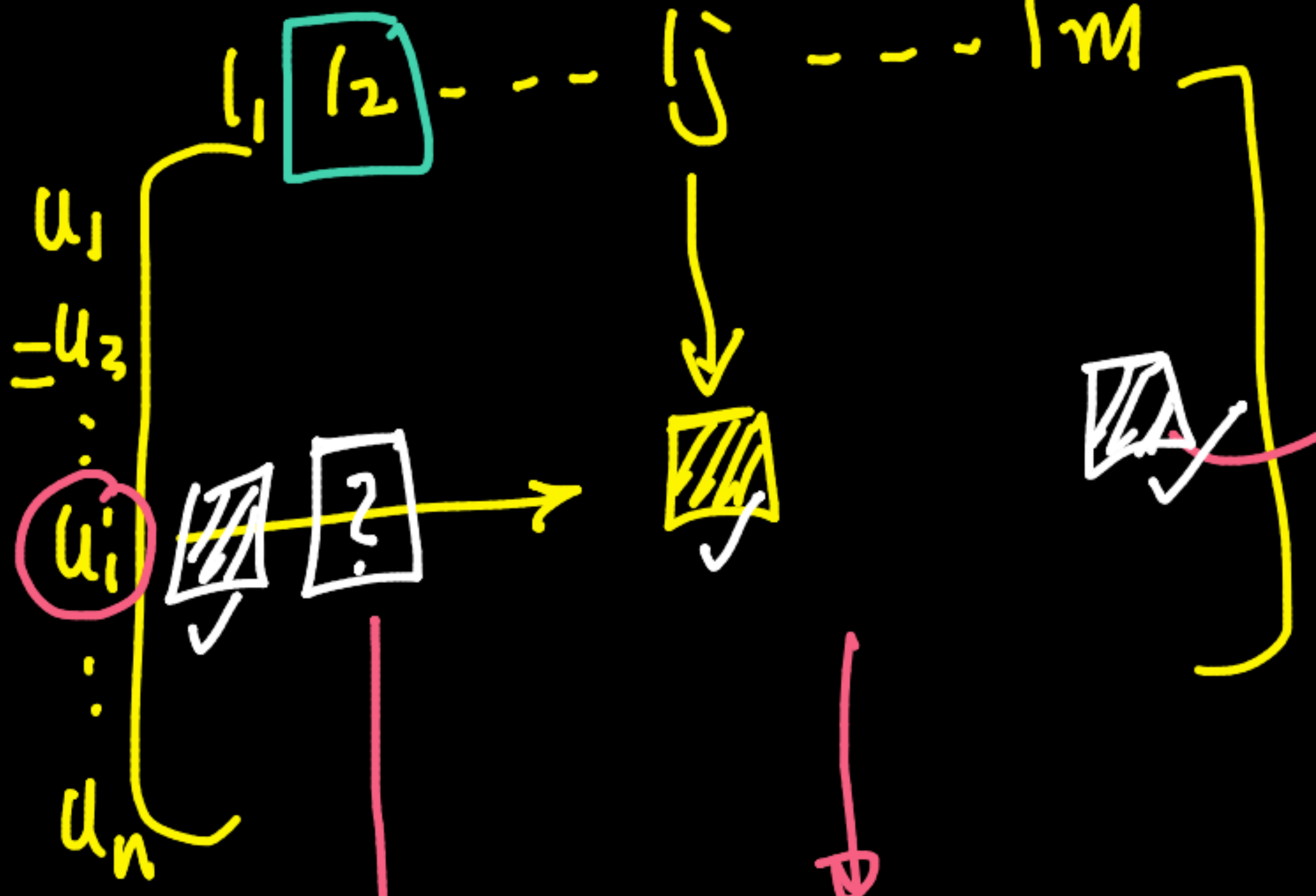
$$A_{n \times m} = \underbrace{B_{n \times d} \cdot C_{d \times m}}_{\text{Factors}}$$

$$A_{n \times m} = \underbrace{B_{n \times d} \; C_{d \times d'} \; D_{d' \times m}}$$

intuition

MF ⟷ RecSys

$A_{n\times m}$



$n \times m$

mostly empty

very sparse

lots of missing value

RecSys $\longleftarrow$ ✓ Matrix Completion $\longleftarrow$ later (?)

MF or decomposition

✓ Matrix Completion

Linear-reg: Linear-function $x_i \xrightarrow{f} y_i$

DT: axes parallel hyperplanes x_i's of diff-classes

Matrix Completion $\longrightarrow$ Tons of methods

Deep-Learning (SOTA)

(later)

$$A_{ij} = B_i \cdot C_j^T$$

ratings

$$\underbrace{A}_{n\times m} = \underbrace{B}_{n\times d} \cdot \underbrace{C^T}_{d\times m}$$

users items

Interaction-model

$C_{m\times d}$

$$\begin{array}{c|c} & 1\ 2\ 3\ \cdots\ d \\ \hline 1 & \\ 2 & \\ 3 & \\ \vdots & \\ j & \\ \vdots & \\ m & \end{array} \quad m\times d$$

C_j^T

Item-vec

$$\begin{array}{c|c} & 1\ 2\ \cdots\cdots\ d \\ \hline 1 & \\ 2 & \\ 3 & \\ \vdots & \\ i & \\ \vdots & \\ n & \end{array} \quad n\times d$$

B_i^T

user-vec

interaction

user i's rating on item j

$$\underset{1\times 1}{A_{ij}} = \overbrace{[B_i]^T_{1\times d} \cdot [C_j]_{d\times 1}}^{\text{interaction}}$$

d-dim vec for u_i

d-dim 4 item j

Multiplicative model

$B_i \in$ d-dim

$C_j \in$ d-dim

2008-9 : Netflix prize

Fundamental assumption of MF-based RecSys → rating : A_{ij} can be decomposed a dot of u_i & l_j vec

Chrome File Edit View History Bookmarks Profiles Tab Window Help

$\rightarrow$ find

$\forall i,j \quad A_{ij} = B_i^T \cdot C_j$

$\forall i: 1 \to n$

$B_i \in \mathbb{R}^d$

$C_j \in \mathbb{R}^d$

$\forall j: 1 \to m$

Some A_{ij}'s in given data

✓ $n = 10{,}000$

✓ $m = 1000$

$n \times m = 10^7$

we only have 10^5 values of A_{ij}'s

11 / 13

$\forall i, j$ s.t. A_{ij} is known/given ; $A_{ij} \approx B_i^T C_j$

$$\min_{B_i,\, C_j} \sum_{\substack{i,j \\ \text{s.t. } A_{ij} \text{ is known}}} \left(\overset{\text{given}}{A_{ij}} - \underbrace{B_i^T \cdot C_j}_{\text{interaction model}} \right)^2$$

n $\quad$ m

Mean-Sq. loss

Solve

1. SGD — B_i's & G_i's

2. Coordinate descent
   - ↳ fix B_i's & update G_i's
   - → fix G_i's & update B_i's

Using only non-empty cells : A_{ij} + SGD/optmzn

$$\begin{cases} B_i & \forall i : 1 \to n \\ C_j & \forall j : 1 \to m \end{cases}$$

(Q) How can I complete the matrix

↳ fill the missing cells

$A_{3,10} \rightarrow$ missing

$A_{ij} \approx B_i^T C_j$

$\left\{ A_{3,10} \approx B_3^T \cdot C_{10} \right\}$ ***

$B_3 \rightarrow A_{3,j}$

$\downarrow$

other ratings by U_3

$C_{10} \rightarrow A_{i,10}$

$\hookrightarrow$ other ratings on I_{10}

✓ $A_{1,10}$ ✓ $A_{2,10}$ ✗ $A_{3,10}$ . . . . ✓ $A_{n,10}$

$$\min_{B_i, C_j} \sum_{i,j} \left(A_{ij} - B_i^{T} C_j \right)^2$$

s.t. $A_{ij} \neq$ NULL

✓ $A_{3,1}$

✗ $A_{3,2}$

✓ $A_{3,3}$

$\vdots$

✗ $A_{3,m}$

vec $\cdot\ I_{10}$

e.g.: $\boxed{C_{10}}$

e.g.: $B_3 \leftarrow$ user$_3$

Failure Case

new-item / cold start

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & \cdots & 150 & \cdots & m \\
\hline
 & & & & \square & & \\
 & & & & \square & & \\
 & & & & \square & & \\
u_{100} & A_{100,1} & A_{100,2} & \cdots & \cdots & \cdots & A_{100,m} \\
 & & & & \vdots & & \\
 & & & & \square & &
\end{array}
$$

$A =$ (matrix above)

new-user

✓ cold start

$$\begin{tikzcd}
\text{RecSys} \arrow[r, leftrightarrow] & \text{Matrix completion} \arrow[d, leftrightarrow] \\
& \boxed{\text{MF}} \arrow[d] \\
& \text{optmzn}
\end{tikzcd}$$

$A_{ij} \approx B_i^T C_j$

MF : PCA ✓

standardized data

$X_{n \times d}$ ✓

$$X_{n\times d} \longrightarrow Cov(X) = S_{d\times d} = X^T_{d\times n} \cdot X_{n\times d}$$

Sq. symm matrix → $\boxed{S_{d\times d}}$

$$S_{d\times d} = W_{d\times d} \cdot \Lambda_{d\times d} \cdot W^T_{d\times d}$$

$$W^T = \begin{bmatrix} \leftarrow & v_1^T & \rightarrow \\ \leftarrow & v_2^T & \rightarrow \\ \leftarrow & v_d^T & \rightarrow \end{bmatrix}$$

$$W = \begin{bmatrix} \uparrow & \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & v_3 & \cdots & v_d \\ \downarrow & \downarrow & \downarrow & & \downarrow \end{bmatrix}_{d\times d}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{bmatrix}$$

diagonal matrix

$$S_{d\times d} = W_{d\times d} \cdot \Lambda_{d\times d} \cdot W^T_{d\times d}$$

each row is $\perp$r to other

each col is $\perp$ other

diagonal matrix

① [PCA = special type of MF] ✓

PCA : eigen-decomposition

$$X^T_{d\times n} X_{n\times d} = S_{d\times d} = W \Lambda W^T$$

② eig vec as its col

eig-val (λ_i's)

③ SVD

data

$X_{n\times d}$

$n>d$

③ $X_{n\times d} = U_{n\times n} \Sigma_{n\times d} V^T_{d\times d}$

singular values

$$s_i = \sqrt{\lambda_i}$$

$$\begin{bmatrix} s_1 & & & 0 & \\ & s_2 & & & \\ & & \ddots & & \\ 0 & & s_d & 0 & 0 \; 0 \; 0 \end{bmatrix}_{n\times d}$$

4

$U_{n\times n}$

$$\begin{bmatrix} \uparrow & \uparrow & & \uparrow & & \uparrow \\ u_1 & u_2 & \ldots & u_i & \ldots & u_n \\ \downarrow & \downarrow & & \downarrow & & \downarrow \end{bmatrix}_{n\times n}$$

U_i = ith eigvec of $\underset{n\times d}{X} \cdot \underset{d\times n}{X^T} = S'_{n\times n}$

5

$V_{d\times d}$

v_i : ith col of V

↳ ith eig vec of

$\underset{d\times n}{X^T}\,\underset{n\times d}{X} = S_{d\times d}$

ith col of W

truncated SVD (later)

$\hookrightarrow X_{n \times d}$: Feature engineering using SVD

MF for Clustering

(K-means)

K-means:

$$\mathcal{D} = \{x_i\}_{i=1}^{n}$$

Given x_i's ; find $c_j : 1 \to k$

$i : 1 \to n$

$$\min_{c_j} \sum_{j=1}^{k} \sum_{x_i \in c_j} \overbrace{\| \underbrace{x_i}_{d\text{-dim}} - \underbrace{c_j}_{d\text{-dim}} \|}^{\text{intra-clust dist}}{}^{2}$$

find

each x_i should belong to c_j

Let's define matrix Z s.t

$$z_{ij} = \begin{cases} 1 & \text{if } x_i \in S_j \\ 0 & \text{o/w} \end{cases}$$

S_j → set of pts in cluster j

Cluster assignment matrix

$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & \cdots & j & \cdots & K \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ n \end{matrix} & \begin{bmatrix} & & & & & \\ & & & & & \\ & & & & & \\ 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 \\ & & & & & \\ & & & & & \end{bmatrix} \end{array}$$

$$\min_{c_j} \sum_{j=1}^{K} \sum_{x_i \in S_j} \|x_i - c_j\|^2 \simeq \min_{c_j,\, z_{ij}} \sum_{i=1}^{n} \overbrace{\underbrace{z_{ij}}_{1 \text{ if } x_i \in S_j}}^{0} \cdot \|x_i - c_j\|^2$$

s.t. $z_{ij} = 1$ or 0

$$\sum_{j=1}^{K} z_{ij} = 1$$

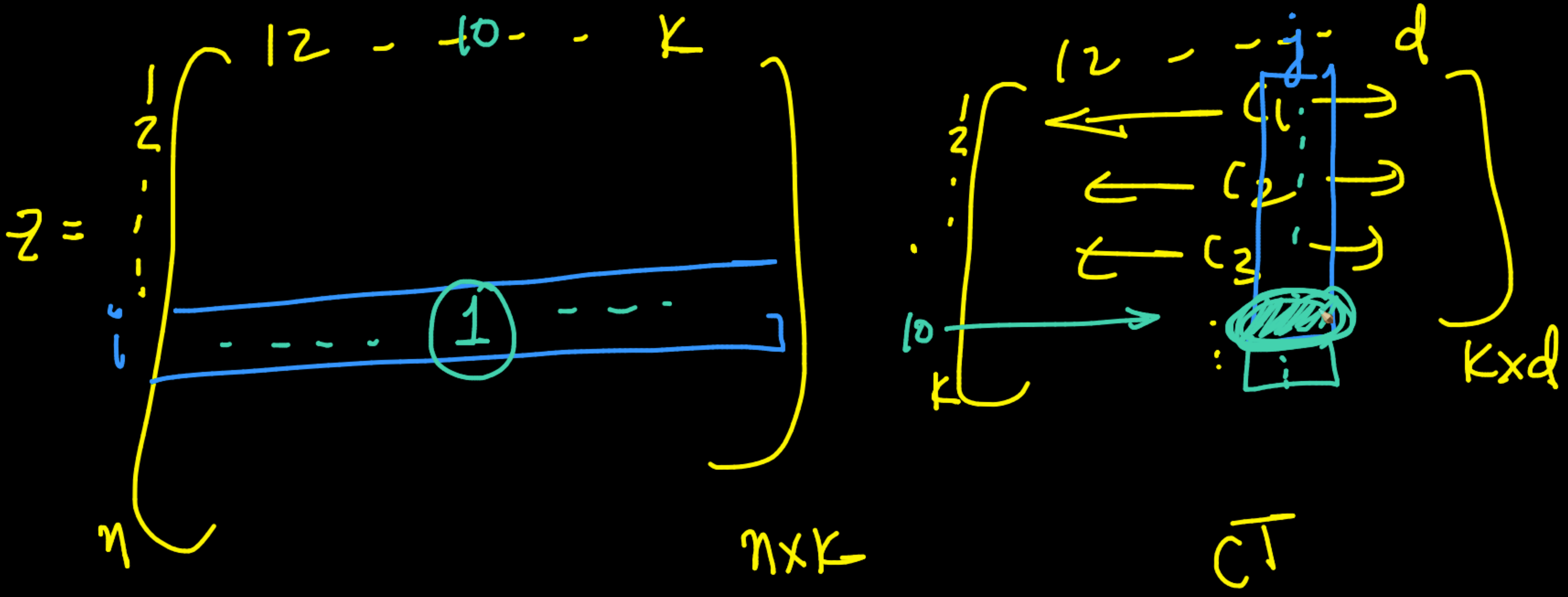
$$Z = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} 1 \quad 2 \quad \cdots \quad K \\ \left[\begin{array}{ccc} & & \\ & \square & \\ & & \end{array} \right] \end{array}_{n \times K}$$

$$C = \begin{bmatrix} \uparrow & \uparrow & & \uparrow & & \uparrow \\ c_1 & c_2 & \cdots & c_j & \cdots & c_K \\ \downarrow & \downarrow & & \downarrow & & \downarrow \end{bmatrix}_{d \times K}$$

$$X^T = \begin{bmatrix} \uparrow & \uparrow & & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_i & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow & & \downarrow \end{bmatrix}_{d \times n}$$

$$X = \begin{bmatrix} \leftarrow & x_1 & \rightarrow \\ \leftarrow & x_2 & \rightarrow \\ & \vdots & \\ \leftarrow & x_n & \rightarrow \end{bmatrix}_{n \times d}$$

j



$i \in S_{10}$

Squared norm : $\|w\|^2 = w_1^2 + w_2^2 + \dots + w_d^2$

K-means Clust

$$X_{n\times d} \approx Z \cdot C^T$$

centroid matrix

Cluster alloc matrix

↳ Constraints

$$A = B^T \cdot C$$

$$\min_{B,C} \|A - B^T \cdot C\|_F^2$$

$$\simeq \min_{B,C} \sum_{i,j} \left(A_{ij} - B_i^T C_j\right)^2$$

K-means

$$\min_{Z,C} \|X - Z \cdot C^T\|_F^2$$

$$\text{s.t. } z_{ij} = 1 \text{ or } 0$$

$$\sum_{j=1}^{K} z_{ij} = 1$$

$$A = B^T \cdot C$$

$$\min_{B,C} \| A - B^T \cdot C \|_F^2$$

$$\simeq \min_{B,C} \sum_{i,j} \left(A_{ij} - B_i^T C_j \right)^2$$

K-means

$$\min_{Z,C} \| \underset{n \times d}{X} - \underset{n \times k}{Z} \cdot \underset{k \times d}{C^T} \|_F^2$$

s.t. $Z_{ij} = 1$ or 0 ✗

$$\sum_{j=1}^{K} Z_{ij} = 1 \quad ✓$$

Takeaway:

MF → completion → RecSys

MF → PCA, SVD

MF → k-means (hard-assign)

MF → GMM (soft-assign)

Chrome File Edit View History Bookmarks Profiles Tab Window Help

$d \rightarrow$ hyper-param

How to find d

(RecSys)

$\updownarrow$

Matrix completion

$$\underset{n\times m}{A} = \underset{n\times d}{B}\ \underset{d\times m}{C^T}$$

One global minima

$$\min_{\substack{B_i, C_j \\ \in \mathbb{R}^d}} \sum_{\substack{i,j \\ \text{s.t. } A_{ij} \neq \text{NULL}}} \left(A_{ij} - B_i^T C_j\right)^2 \approx \min_{B,C} \|A - BC^T\|_F^2$$

36 / 39

$$A_{i_0,j} \longleftarrow B_{i_0} \in \mathbb{R}^d$$

$$A_{i,j_0} \longleftarrow C_{j_0} \in \mathbb{R}^d$$

Chrome File Edit View History Bookmarks Profiles Tab Window Help

option 2 hyper-param tune ?

better

Some Aij (80%) random that are not NULL

↓

Train (B, C)

20% → Test Set

38 / 38


Chrome File Edit View History Bookmarks Profiles Tab Window Help
TOI/TH
1B+
Ui
Wj
200 M+
Facebook
Google
Amazon
advertisers
(Nestle)
U1
U2
U3
Ui
Un
100
W1
W2
Wj
Wm
40 / 42

$$A_{n \times m} = \begin{array}{cc} & w_j \\ u_i & \begin{bmatrix} \quad & \quad & \quad \\ & \boxed{100} & \\ & & \end{bmatrix} \end{array}$$

$$A_{n \times m} = U_{n \times d} \, W^T_{d \times m}$$

do NOT any meaning

$$U = \begin{bmatrix} \leftarrow & u_1^T & \rightarrow \\ \leftarrow & u_2^T & \rightarrow \\ \leftarrow & \boxed{u_i^T} & \rightarrow \\ \leftarrow & u_n^T & \rightarrow \end{bmatrix}$$

$n \times d$

lin

d-dim rep of user i

hyper param

$\left\{ \text{Sim}(u_i, u_j) \text{ is high} \right.$

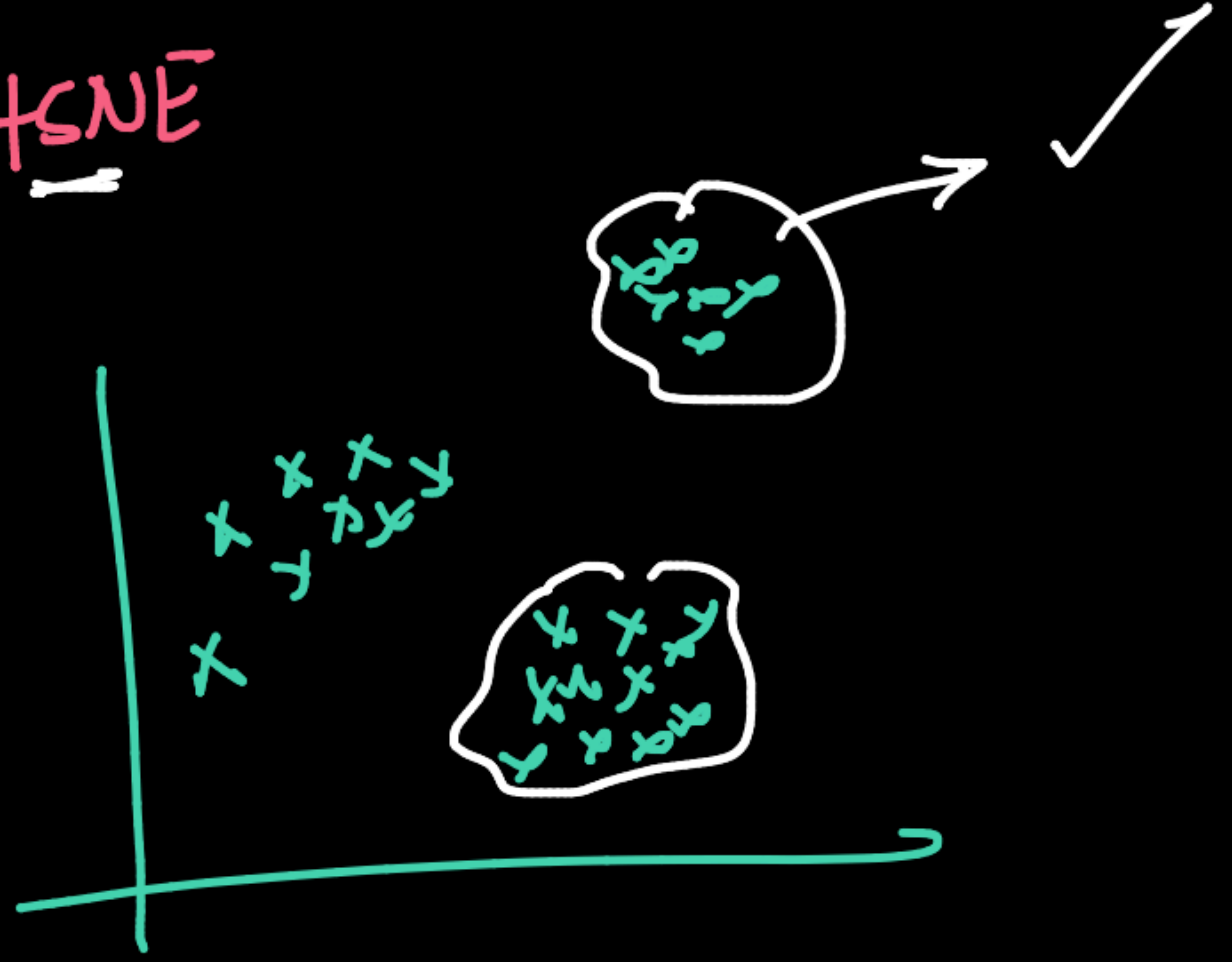
very similar browsing

Interpretability

$u_i \quad \forall i : 1 \to n \; ; u_i \in \mathbb{R}^d$

$\downarrow$ UMAP | tSNE

$u_i' \in \mathbb{R}^2$



Chrome File Edit View History Bookmarks Profiles Tab Window Help

11:18

- NMF (5 min)
- Netflix Prize (30 min)
- MF for featuse engg (next class)

44 / 44

$$\min_{B,C} \|A - B^T C\|_F^2$$

Rec Sys

$A_{ij} \approx B_i^T C_j$: mean SQ. loss min

MF

PCA: $S_{d\times d} = X^T X = W \Lambda W^T$

SVD: $X_{n\times d} = U_{n\times n} \Sigma_{n\times d} V^T_{d\times d}$

eigvec of XX^T

$s_i = \sqrt{\lambda_i}$

d?

elbow

Train Test hyper-param

Clustering

$$\min_{Z,C} \|X - ZC^T\|_F^2$$

Z : constraints

Q

$X_{n\times d}$

① $X^T_{d\times n} X_{n\times d} = S_{d\times d} \rightarrow$ d eig-values; eig-vec are d-dim

② $X_{n\times d} X^T_{d\times n} = S'_{n\times n} \rightarrow$ n eig-val; eig-vec are n-dim

$S \neq S'$

Math

B_i : col vec (default)

$$A_{ij} = B_i^T C_j \simeq A_{n\times m} = B_{n\times d}\, C^T_{d\times m}$$

$$B = \begin{array}{c|c} & 1\ 2\ \cdots\ d \\ \hline \begin{array}{c}1\\2\\\vdots\\i\\ \\n\end{array} & \left[\begin{array}{c} \\ \\ \\ \longleftarrow \underline{B_i} \longrightarrow \\ \\ \\ \end{array}\right] \end{array}$$

$$\underset{n\times d}{C} = \begin{array}{c|c} & 1\ 2\ \cdots\ d \\ \hline \begin{array}{c}1\\2\\\vdots\\j\\\vdots\\m\end{array} & \left[\begin{array}{c} \\ \\ \\ \longleftarrow \underline{C_j} \longrightarrow \\ \\ \\ \end{array}\right] \end{array}$$


Chrome File Edit View History Bookmarks Profiles Tab Window Help
11:36
New Tab
Search Google or type a URL
Update
Gmail
Images
Google
Search Google or type a URL
Colaboratory
YouTube
My Drive
InterviewBit S...
Learning
GitHub
Reduce the fil...
Scaler Acade...
(99+) Feed
Add shortcut
49 / 49
Photo by NASA Image Library


Chrome
File
Edit
View
History
Bookmarks
Profiles
Tab
Window
Help
11:36
New Tab
Search Google or type a URL
Update
Gmail
Images
Google
Search Google or type a URL
Colaboratory
YouTube
My Drive
InterviewBit S...
Learning
GitHub
Reduce the fil...
Scaler Acade...
(99+) Feed
Add shortcut
49 / 49
Photo by NASA Image Library