# Matrix factorizations

Vladislav Goncharenko

Fall 2023, Harbour Space

Materials provided by Dzen and Andrey Zimovnov

# Types of problem statements recommendations

In general, we want to recommend items (music/movie/article/product) to the user so that the user is happy (we work on long-term metrics)

A typical simplification of a user's happiness is whether a recommendation is liked at the moment. Or even a bigger simplification —do you like the item as a whole or any item rating for the user

The user himself can also be represented in different ways

1. As an unordered set of items that the user interacted with
2. As an unordered set of items with ratings
3. As a sequence of items with ratings
4. As a sequence of interactions taking into account the context

# Revise

girafe
ai

01

Basically, we will represent the user as an unordered set of items with ratings and try to predict the user's rating on other items

# Evaluation matrix

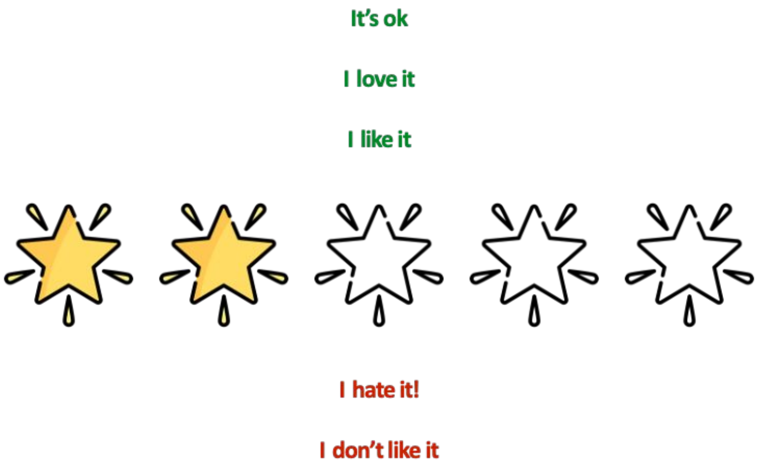
Basically, we will represent the user as an unordered set of items with ratings and try to predict the user's rating on other items

Movie


Users

| | | | | | |
|---|---|---|---|---|---|
| 2 | | 2 | 4 | 5 | |
| 5 | | 4 | | | 1 |
| | | 5 | | 2 | |
| | 1 | | 5 | | 4 |
| | | 4 | | | 2 |
| 4 | 5 | | 1 | | |

# Types of ratings (feedback)




Explicit (from 1 to 5)

Implicit (have you watched this movie)

# Item- and User-based approaches

girafe
ai

02

# Idea

## User2User


| | | | | | |
|---|---|---|---|---|---|
| 2 | | 2 | 4 | 5 | |
| 5 | | 4 | | | 1 |
| | | 5 | | 2 | |
| | 1 | | 5 | | 4 |
| | | 4 | | | 2 |
| 4 | 5 | | 1 | | |

## Item2Item


| | | | | | |
|---|---|---|---|---|---|
| 2 | | 2 | 4 | 5 | |
| 5 | | 4 | | | 1 |
| | | 5 | | 2 | |
| | 1 | | 5 | | 4 |
| | | 4 | | | 2 |
| 4 | 5 | | 1 | | |

# A few formulas

$$\hat{r}_{ui} = \frac{\sum_j s(i,j)\left(r_{uj} - r_u\right)}{\sum_j |s(i,j)|} + r_u$$

Explicit:

$$corr(i,j) = \frac{\sum_u (r_{ui} - r_i)(r_{uj} - r_j)}{\sqrt{\sum_u (r_{ui} - r_i)^2 \sum_u \left(r_{uj} - r_j\right)^2}}$$

$s(i,j)$

Implicit:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

# Choosing between Item2Item and User2User

It strongly depends on the number of ratings per user and per item – the more, the more reliable the similarity.

This also leads to the possibility of pre-calculation: if there are a lot of estimates, then adding a pair of estimates will not change anything. Then you can update once in a while.

The similarities themselves can be used for other tasks (candidate generation, contextual recommendations).

# A bit of linear algebra

girafe
ai

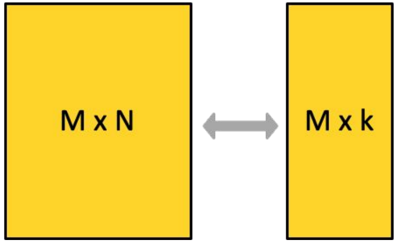
03

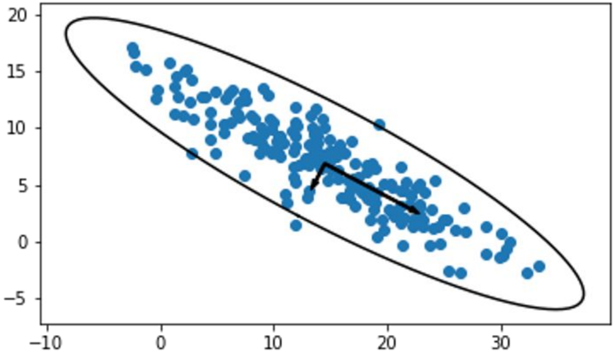
# Recall the main things about

- PCA
- SVD
- Truncated SVD

# PCA – staging

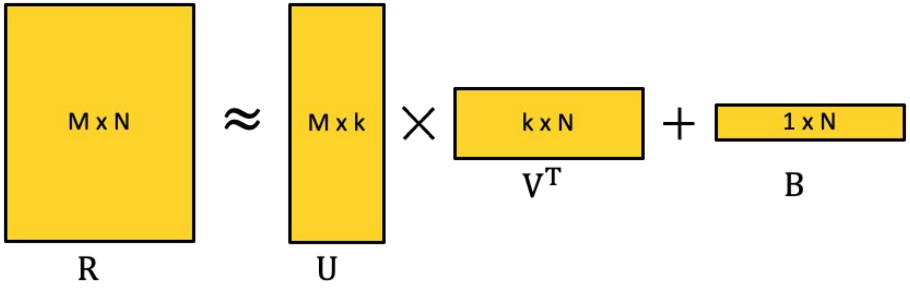
$$\sum_i dist(x_i, L_k) \to \min$$

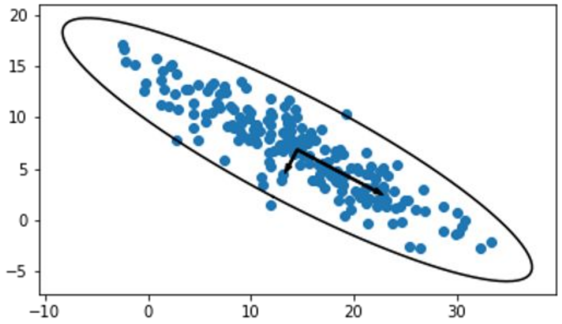


# PCA – connection with matrix decompositions

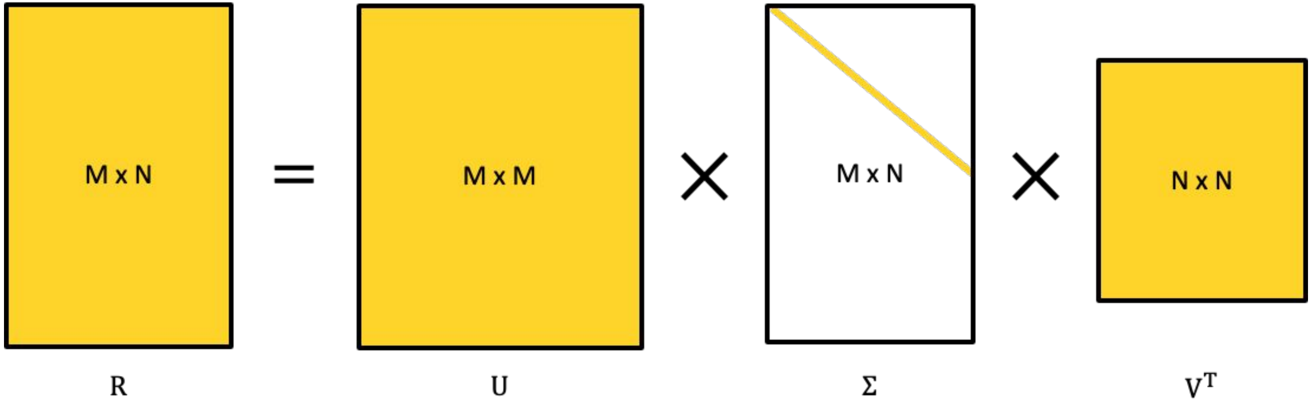
$$r_{ij} = \overline{u_i}^T \overline{v_j} + b_i$$



# SVD

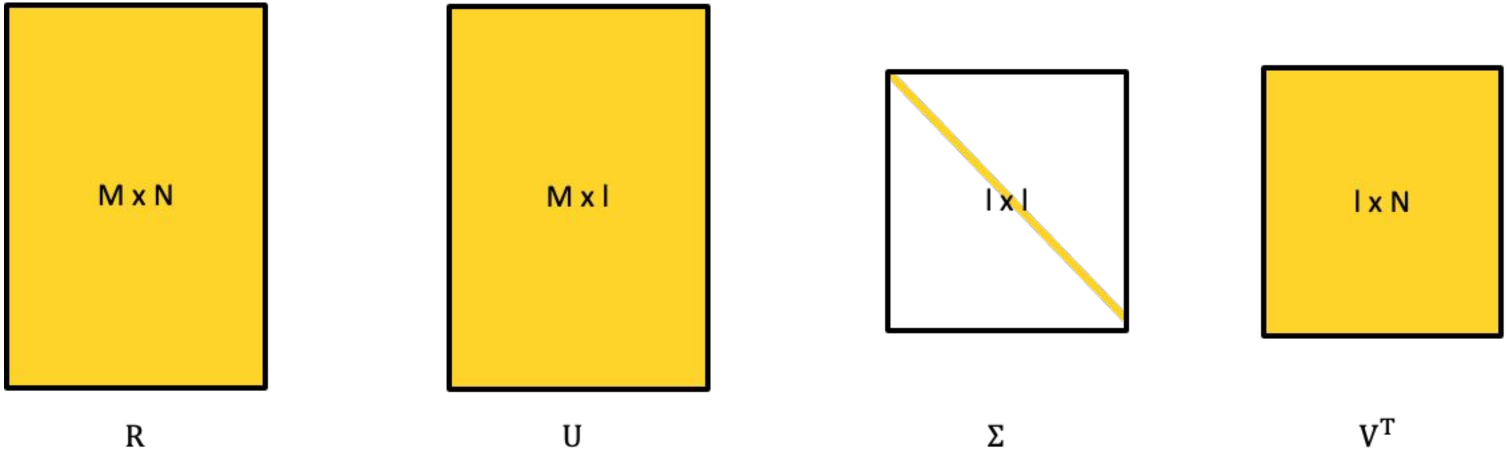
$$R = U\Sigma V^T$$
$$UU^T = U^TU = I_M$$
$$VV^T = V^TV = I_N$$




$$r_{ij} = \bar{u_i}^T \Sigma \bar{v_j}$$

# SVD compact

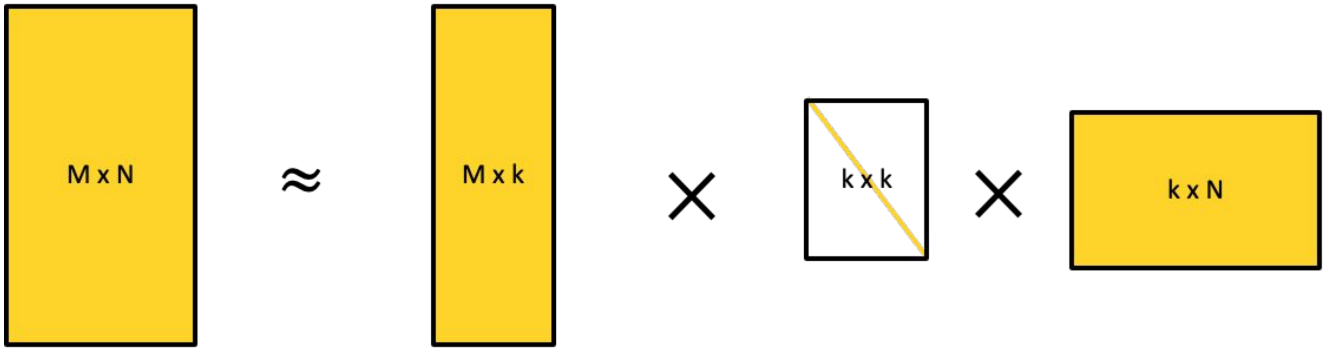
$$R = U\Sigma V^T$$
$$l = rank\{R\}$$
$$U^T U = V^T V = I_l$$




$$r_{ij} = \bar{u_i}^T \Sigma \bar{v_j}$$

# Truncated SVD

$$R \approx U\Sigma \mathrm{V}^{\mathrm{T}}$$
$$k \ll l = \operatorname{rank}\{R\}$$
$$U^T U = V^T V = I_k$$




$$r_{ij} = \bar{u_i}^T \Sigma \bar{v_j}$$

# Truncated SVD as optimization

$$R \approx U\Sigma V^{T}$$
$$U^{T}U = V^{T}V = I_k$$

$$U\Sigma V^{T} = XY^{T}$$
$$X^{T}X \neq I_k, Y^{T}Y \neq I_k$$

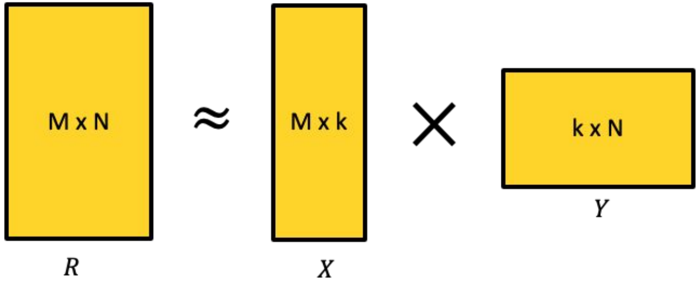
$$\min_{\substack{U,V,\Sigma \\ U^TU=V^TV=I_k \\ \sigma_{ij}=0, i\neq j \\ \sigma_{ii}>\sigma_{jj}, i<j}} \sum_{\forall i,j} \left(r_{ij} - u_i^T \Sigma v_j\right)^2 \Leftrightarrow \min_{X,Y} \sum_{\forall i,j} \left(r_{ij} - x_i^T y_j\right)^2$$

$$\min_{X,Y} \sum_{\forall i,j} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \|y_j\|^2 C_j$$

$$C_i = \frac{\left|\{j | r_{ij} > 0\}\right|^{\alpha} |\{i\}|}{\sum_i \left|\{j | r_{ij} > 0\}\right|^{\alpha}}$$

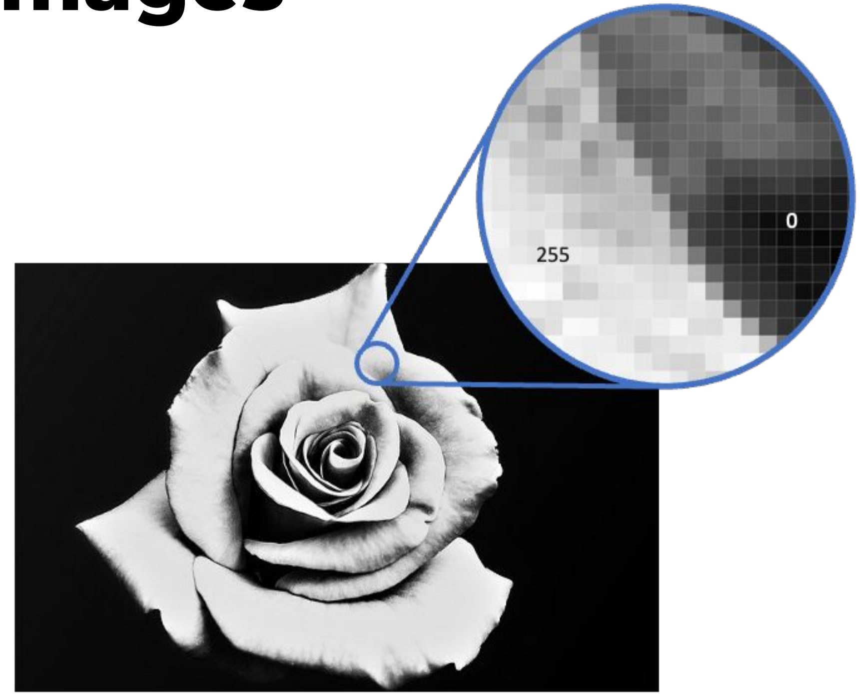
# Truncated SVD as optimization

$$\min_{X,Y} \sum_{\forall i,j} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_{i} \|x_i\|^2 c_i + \lambda \sum_{j} \left\|y_j\right\|^2 c_j$$

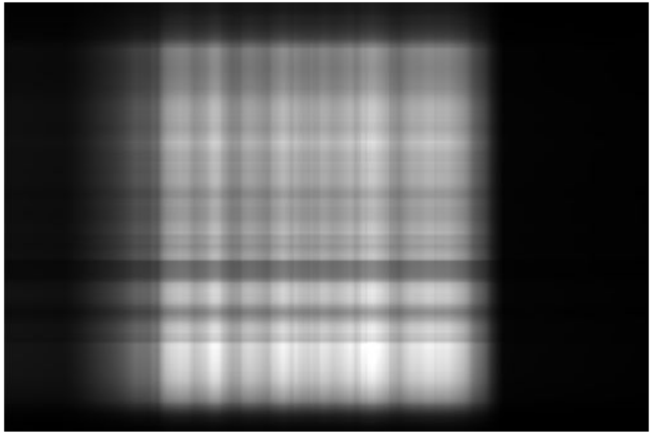



$$r_{i,j} = x_i^T y_j$$

# Truncated SVD on the example of images



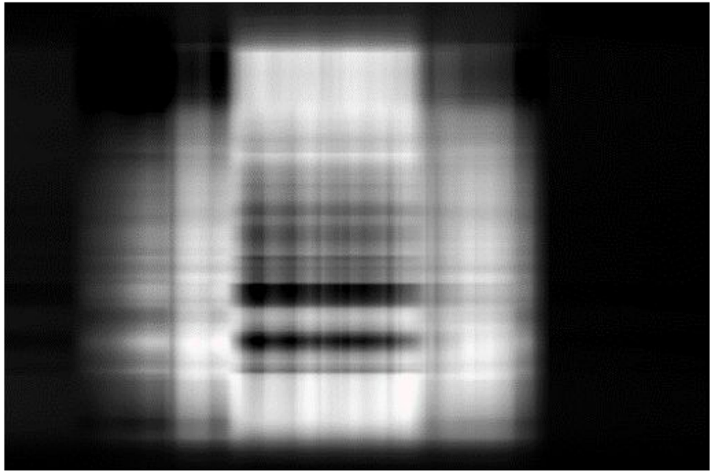

400 x 600 image
240 000 values



400 x 1 / 1 x 1 / 1 x 600 matrices
1001 values

# Truncated SVD on the example of images

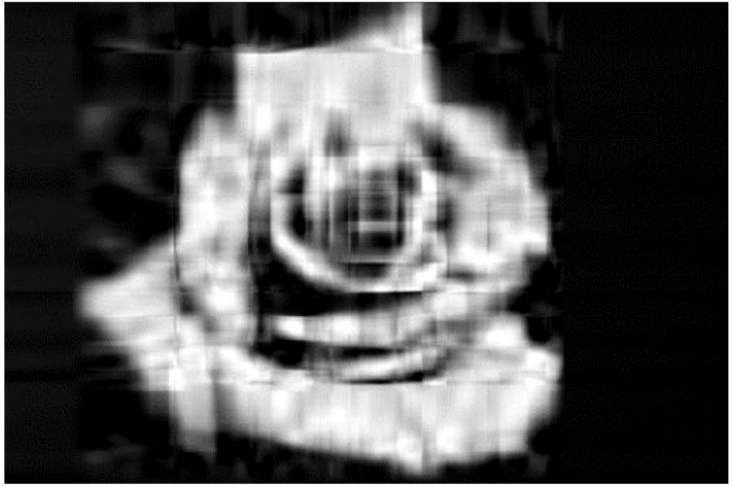
400 x 600 image
240 000 values



400 x 2 / 2 x 2 / 2 x 600 matrices
2002 values

# Truncated SVD on the example of images

400 x 600 image
240 000 values



400 x 9 / 9 x 9 / 9 x 600 matrices
9009 values

# Matrix factorizations

girafe
ai

04

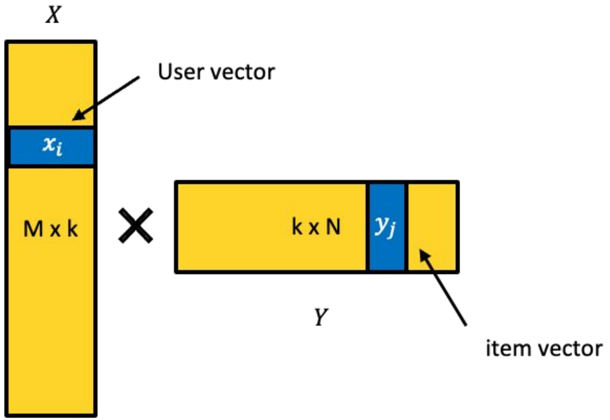
# Funk SVD

Item


Users

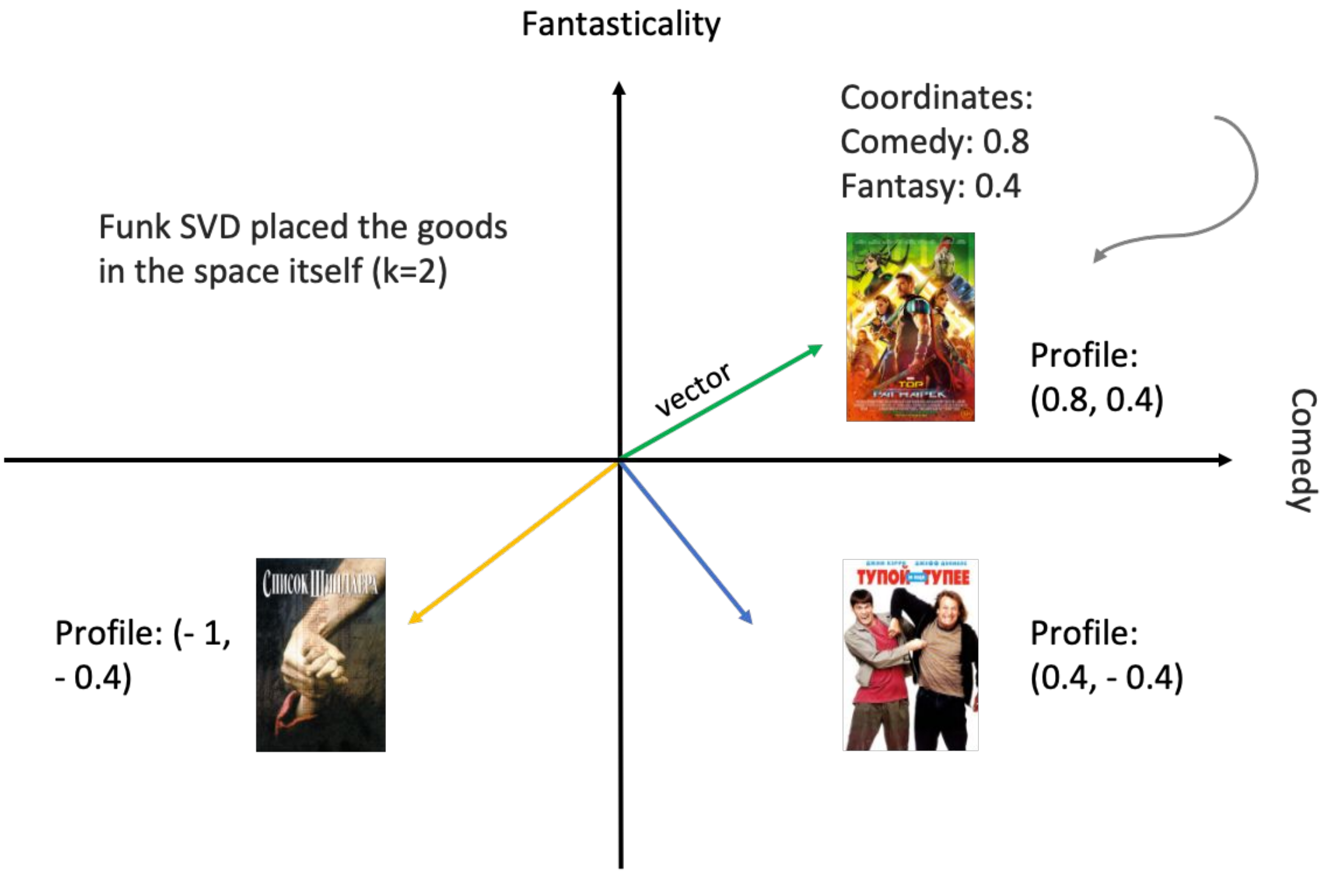
| | | | | | |
|---|---|---|---|---|---|
| 2 | | 2 | 4 | 5 | |
| 5 | | 4 | | | 1 |
| | | 5 | | 2 | |
| | 1 | | 5 | | 4 |
| | | 4 | | | 2 |
| 4 | 5 | | 1 | | |

$\approx$

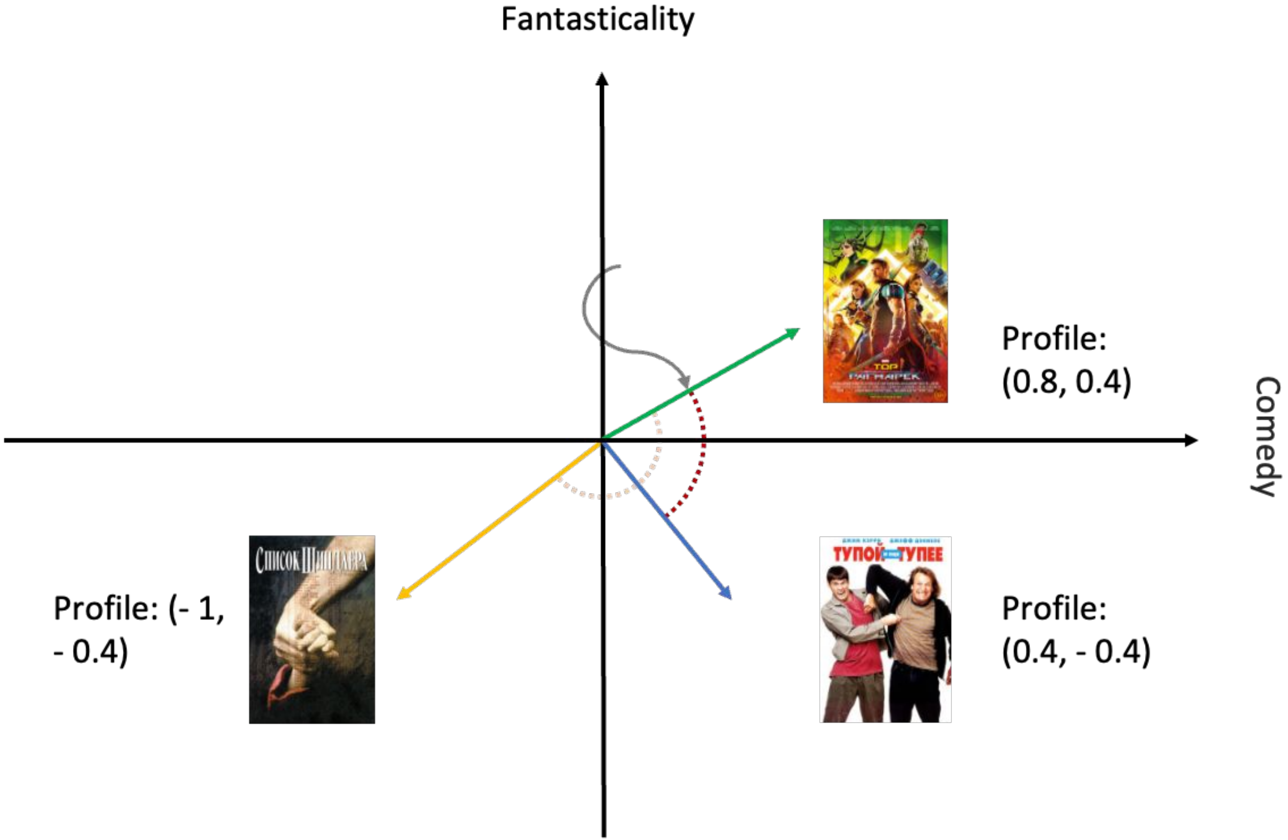



$$\min_{X,Y} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \|y_j\|^2 C_j$$

# SVD Interpretation

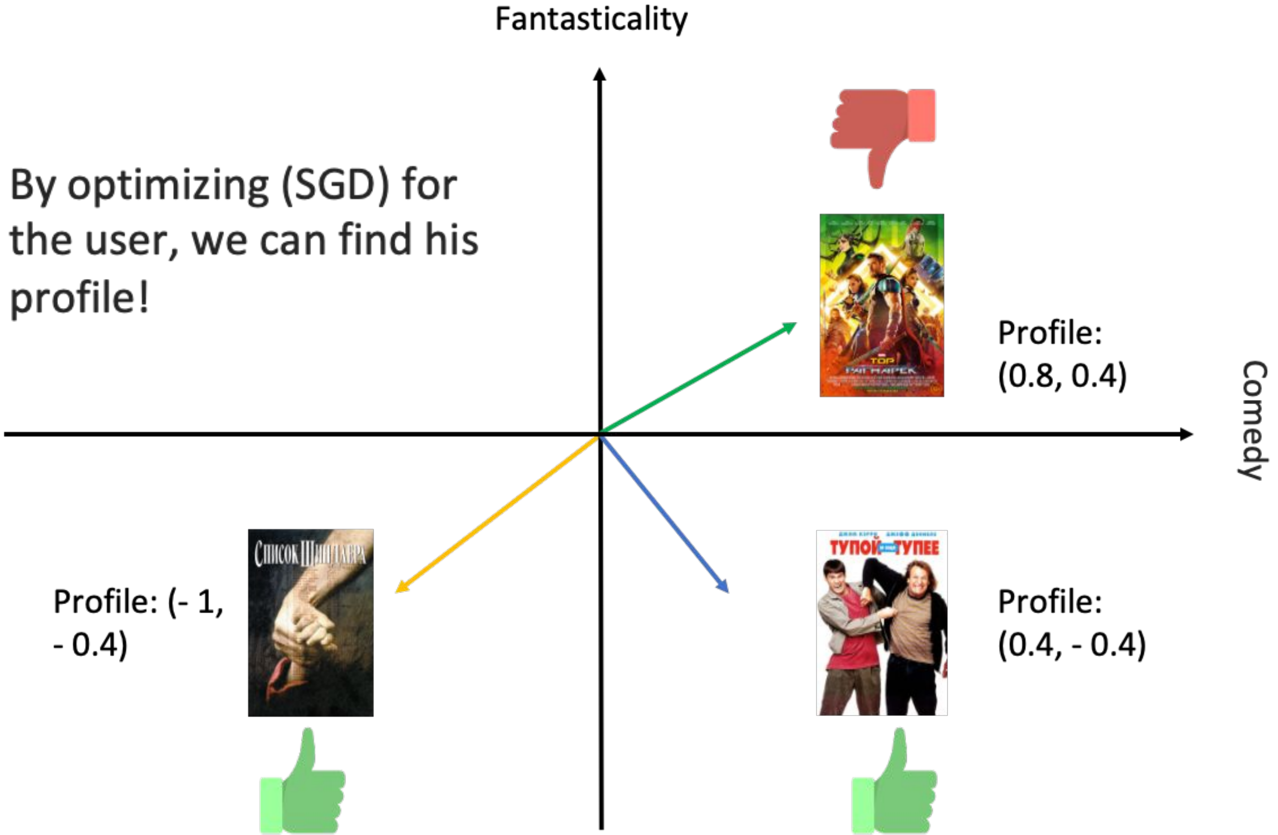


# SVD Interpretation

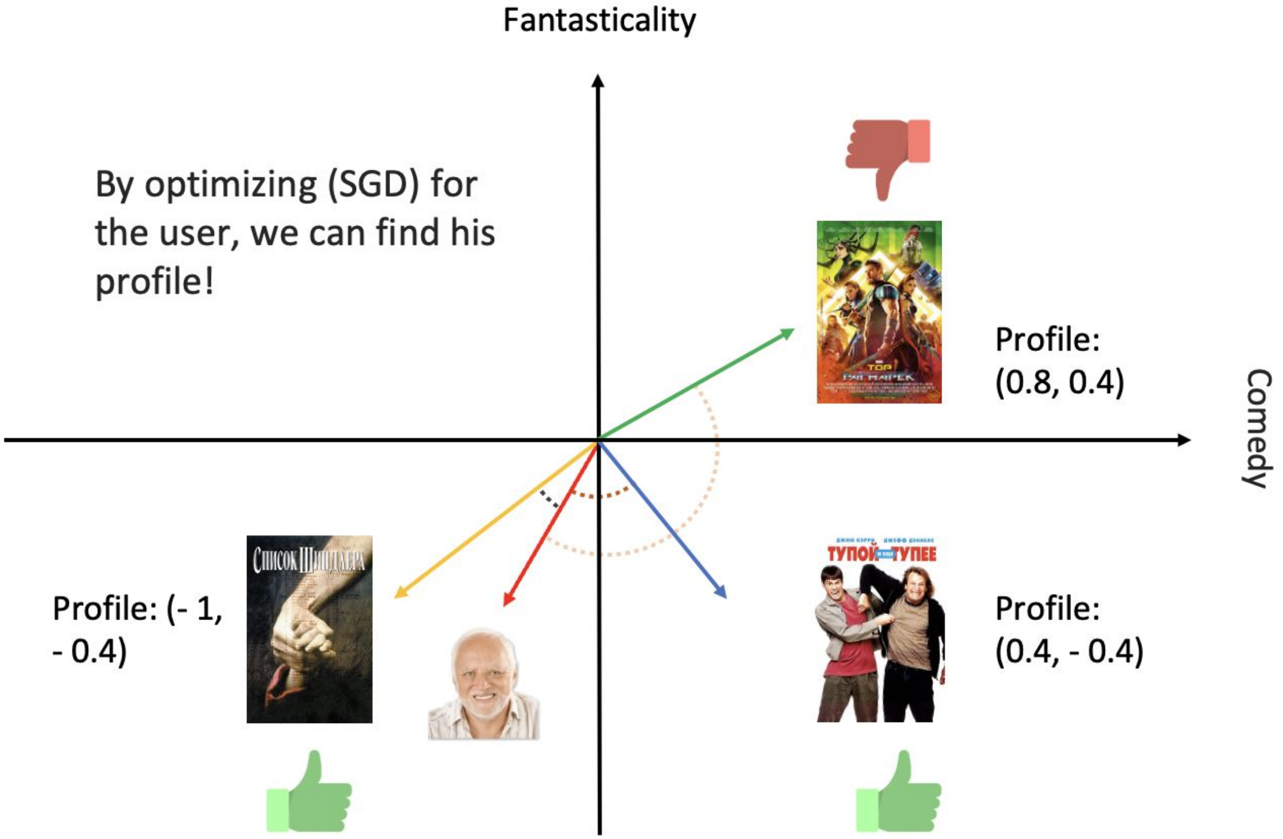


# SVD Interpretation

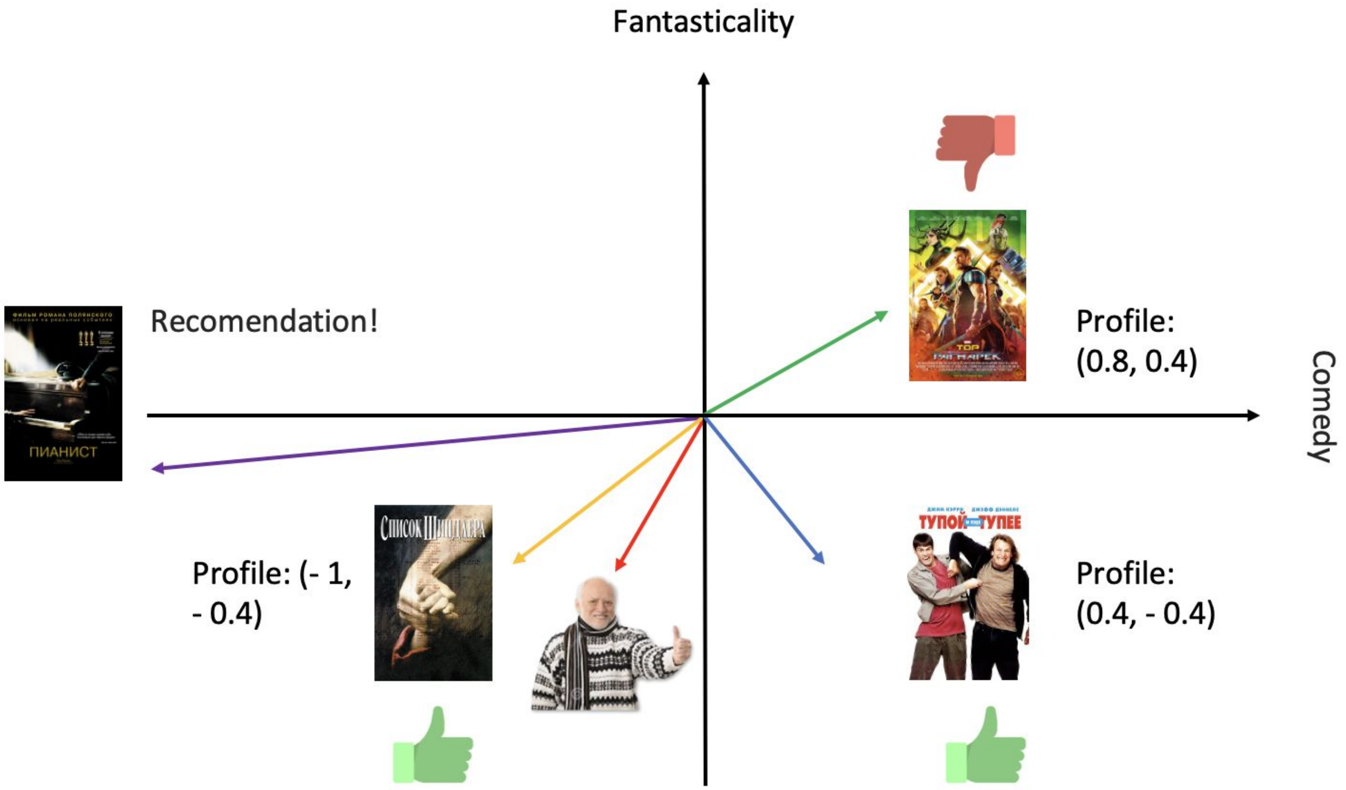
By optimizing (SGD) for the user, we can find his profile!



# SVD Interpretation



# SVD Interpretation



# SVD Results

Matrix factorizations (MF) do not read the similarity of goods directly (as in CF).

MF attempts to describe users and products with a small set of characteristics that explain the ratings.

These characteristics may not be interpreted, at least the winds of items and users are determined with precision to the turn.

Unfortunately, SGD is difficult to parallelize.

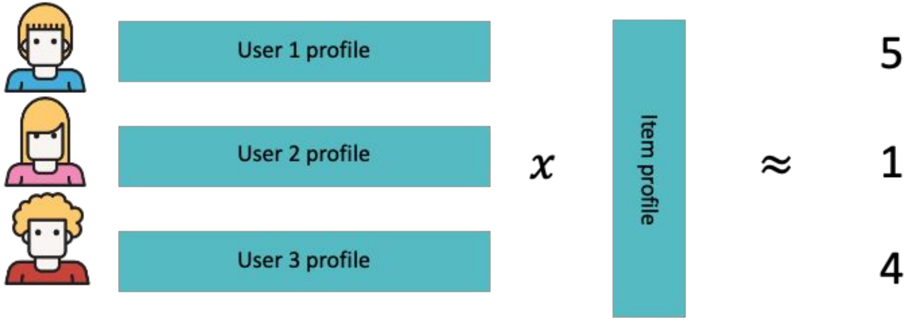
# ALS

$$\min_{X,Y} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \|y_j\|^2 C_j$$

Initialize X and Y with random values.

In the loop:

- We fix the matrix X (users)
- Find the optimal matrix Y (solve the ridge regression for each product)




- And vice versa

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \cancel{\lambda \sum_j \|y_j\|^2 C_j} =$$

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \cancel{\lambda \sum_j \|y_j\|^2 C_j} =$$

$$\underset{x_i}{\operatorname{argmin}} \cancel{\sum_{(i,j)\in R} r_{ij}^2} - 2 \sum_{(i,j)\in R} r_{ij} x_i^T y_j + \sum_{(i,j)\in R} \left(x_i^T y_j\right)^2 + \lambda (x_i, x_i) C_i =$$

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \cancel{\lambda \sum_j \left\|y_j\right\|^2 C_j} =$$

$$\underset{x_i}{\operatorname{argmin}} \cancel{\sum_{(i,j)\in R} r_{ij}^2} - 2 \sum_{(i,j)\in R} r_{ij} x_i^T y_j + \sum_{(i,j)\in R} \left(x_i^T y_j\right)^2 + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot x_i^T y_j + \lambda (x_i, x_i) C_i =$$

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \cancel{\lambda \sum_j \|y_j\|^2 C_j} =$$

$$\underset{x_i}{\operatorname{argmin}} \cancel{\sum_{(i,j)\in R} r_{ij}^2} - 2 \sum_{(i,j)\in R} r_{ij} x_i^T y_j + \sum_{(i,j)\in R} \left(x_i^T y_j\right)^2 + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2 x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot x_i^T\, y_j + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2 x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot y_j^T\, x_i + \lambda C_i x_i^T x_i =$$

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \cancel{\sum_j \|y_j\|^2} C_j =$$

$$\underset{x_i}{\operatorname{argmin}} \cancel{\sum_{(i,j)\in R} r_{ij}^2} - 2 \sum_{(i,j)\in R} r_{ij} x_i^T y_j + \sum_{(i,j)\in R} \left(x_i^T y_j\right)^2 + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot x_i^T y_j + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot y_j^T x_i + \lambda C_i x_i^T x_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \left(\sum_{(i,j)\in R} r_{ij} y_j\right) + x_i^T \left(\sum_{(i,j)\in R} y_j y_j^T + \lambda C_i\right) x_i =$$

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \cancel{\lambda \sum_j \|y_j\|^2 C_j} =$$

$$\underset{x_i}{\operatorname{argmin}} \cancel{\sum_{(i,j)\in R} r_{ij}^2} - 2 \sum_{(i,j)\in R} r_{ij} x_i^T y_j + \sum_{(i,j)\in R} \left(x_i^T y_j\right)^2 + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot x_i^T\, y_j + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot y_j^T\, x_i + \lambda C_i x_i^T x_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \left( \sum_{(i,j)\in R} r_{ij} y_j \right) + x_i^T \left( \sum_{(i,j)\in R} y_j y_j^T + \lambda C_i \right) x_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T B_i + x_i^T A_i x_i = A_i^{-1} B_i$$

# ALS – step by $x_i$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \cancel{\lambda \sum_j \|y_j\|^2 C_j} =$$

$$\underset{x_i}{\operatorname{argmin}} \cancel{\sum_{(i,j)\in R} r_{ij}^2} - 2 \sum_{(i,j)\in R} r_{ij} x_i^T y_j + \sum_{(i,j)\in R} \left(x_i^T y_j\right)^2 + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot x_i^T y_j + \lambda (x_i, x_i) C_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \sum_{(i,j)\in R} r_{ij} y_j + \sum_{(i,j)\in R} x_i^T y_j \cdot y_j^T x_i + \lambda C_i x_i^T x_i =$$

$$\underset{x_i}{\operatorname{argmin}} -2x_i^T \left(\sum_{(i,j)\in R} r_{ij} y_j\right) + x_i^T \left(\sum_{(i,j)\in R} y_j y_j^T + \lambda C_i\right) x_i =$$

$$\boxed{\left(\sum_{j|(i,j)\in R} \boldsymbol{y_j y_j^T} + \boldsymbol{\lambda C_i I}\right)^{-1} \left(\sum_{j|(i,j)\in R} \boldsymbol{r_{ij} y_j}\right)}$$

# IALS

$$p_{ij} = \begin{cases} 1, & r_{ij} > 0 \\ 0, & r_{ij} \leq 0 \; or \; r_{ij} - undefined \end{cases}$$

**Did you like it?**

# IALS

$$p_{ij} = \begin{cases} 1, & r_{ij} > 0 \\ 0, & r_{ij} \leq 0 \; or \; r_{ij} - undefined \end{cases}$$

Do you like it?

$$c_{ij} = 1 + \alpha \left| r_{ij} \right|$$

How confident are you in $p_{ij}$

# IALS

$$p_{ij} = \begin{cases} 1, & r_{ij} > 0 \\ 0, & r_{ij} \leq 0 \; or \; r_{ij} - undefined \end{cases}$$

**Do you like it?**

$$c_{ij} = 1 + \alpha |r_{ij}|$$

**How confident are you in $p_{ij}$**

$$\min_{X,Y} \sum_{\forall i,j} c_{ij} (p_{ij} - x_i^T y_j)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \|y_j\|^2 C_j$$

$$C_i = \frac{\left(\sum_{j|(i,j)\in R} c_{ij}\right)^{\alpha} |\{i\}|}{\sum_i \left(\sum_{j|(i,j)\in R} c_{ij}\right)^{\alpha}}$$

# IALS: How to optimize

$$\underset{x_i}{\operatorname{argmin}} \sum_{\forall i,j} c_{ij}\left(p_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \left\|y_j\right\|^2 C_j =$$

$$\left(\sum_{\forall j} c_{ij} y_j y_j^T + \lambda C_i I\right)^{-1} \left(\sum_{\forall j} c_{ij} p_{ij} y_j\right) =$$

# IALS: How to optimize

$$\underset{x_i}{\operatorname{argmin}} \sum_{\forall i,j} c_{ij}\left(p_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \left\|y_j\right\|^2 C_j =$$

$$\left(\sum_{\forall j} c_{ij} y_j y_j^T + \lambda C_i I\right)^{-1} \left(\sum_{\forall j} c_{ij} p_{ij} y_j\right) =$$

$$\left(\sum_{\forall j | p_{ij}=0} c_{ij} y_j \cdot y_j^T + \sum_{\forall j | p_{ij} \neq 0} c_{ij} y_j y_j^T + \lambda C_i I\right)^{-1} \left(\sum_{\forall j | p_{ij} \neq 0} c_{ij} p_{ij} y_j\right) =$$

$$\left(\sum_{\forall j} y_j \cdot y_j^T - \sum_{\forall j | p_{ij} \neq 0} y_j \cdot y_j^T + \sum_{\forall j | p_{ij} \neq 0} c_{ij} y_j y_j^T + \lambda C_i I\right)^{-1} \left(\sum_{\forall j | p_{ij} \neq 0} c_{ij} p_{ij} y_j\right) =$$

$$\left(Y^T Y + \lambda C_i I + \sum_{\forall j | p_{ij} \neq 0} (c_{ij} - 1) y_j y_j^T\right)^{-1} \left(\sum_{\forall j | p_{ij} \neq 0} c_{ij} p_{ij} y_j\right)$$

# IALS – generalizations

The target $p_{ij}$ can be not only 1 where there is a signal (we leave the implicit one equal to 0).

In confidence $c_{ij}$ it is not necessary to use 1 as the default, any peer-to-peer matrix.three times.

The model can be slightly complicated by analogy with PCA:

$$\boldsymbol{r_{ij}} \approx \boldsymbol{x_i y_j} \longrightarrow \boldsymbol{r_{ij}} \approx \boldsymbol{x_i y_j} + \boldsymbol{b_i} + \boldsymbol{b_j} + \boldsymbol{\mu}$$

You can combine several types of assessments with each other.

# Other generalizations

ALS

$$r_{ij} \approx x_i y_j + b_i + b_j + \mu$$

SVD++

$$r_{ij} \approx \left( x_i + \frac{1}{\sqrt{|\{s|p_{is} \neq 0\}|}} \sum_{\forall s|p_{is} \neq 0} \widehat{y_s} \right) y_j + b_i + b_j + \mu$$

timeSVD++

$$r_{ij}(t) \approx \left( x_i(t) + \frac{1}{\sqrt{|\{s|p_{is} \neq 0\}|}} \sum_{\forall s|p_{is} \neq 0} \widehat{y_s} \right) y_j + b_i(t) + b_j(t) + \mu$$

# Other generalizations (at the next lecture)

- SLIM
- Factorization machine
- DSSM
- Other losses

# Thanks for attention!

Questions?

girafe
ai