Матричные факторизации

RecSys

lecturer: Mollaev D. E.
Sber AI Lab

План лекции:

- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- Truncated SVD
- Funk SVD
- Интерпретация SVD
- ALS, IALS

План лекции:

- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- Truncated SVD
- Funk SVD
- Интерпретация SVD
- ALS, IALS

Виды постановок задачи рекомендаций

Наша цель — рекомендовать пользователю такие айтемы, которые сделают его счастливым.
Обычно под «счастьем» упрощённо понимают то, понравилась ли рекомендация в текущий момент. Иногда — ещё проще: нравится ли сам айтем или какую оценку пользователь ему поставит.

Пользователя можно моделировать по-разному:

- Как неупорядоченное множество айтемов, с которыми он взаимодействовал
- Как неупорядоченное множество айтемов с привязанными оценками
  Как последовательность айтемов с оценками
- Как последовательность взаимодействий с учётом контекста

Recap: Item2item, User2user

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		

Recap: Item2item, User2user

- **Зависимость от количества оценок**

  Выбор между методами сильно зависит от числа оценок на пользователя и на объект: чем больше оценок, тем надёжнее определяется схожесть между ними.

- **Возможность периодического обновления**

  При большом количестве оценок добавление нескольких новых практически не влияет на результат, поэтому схожести можно пересчитывать с определённой периодичностью, а не в реальном времени..

Recap

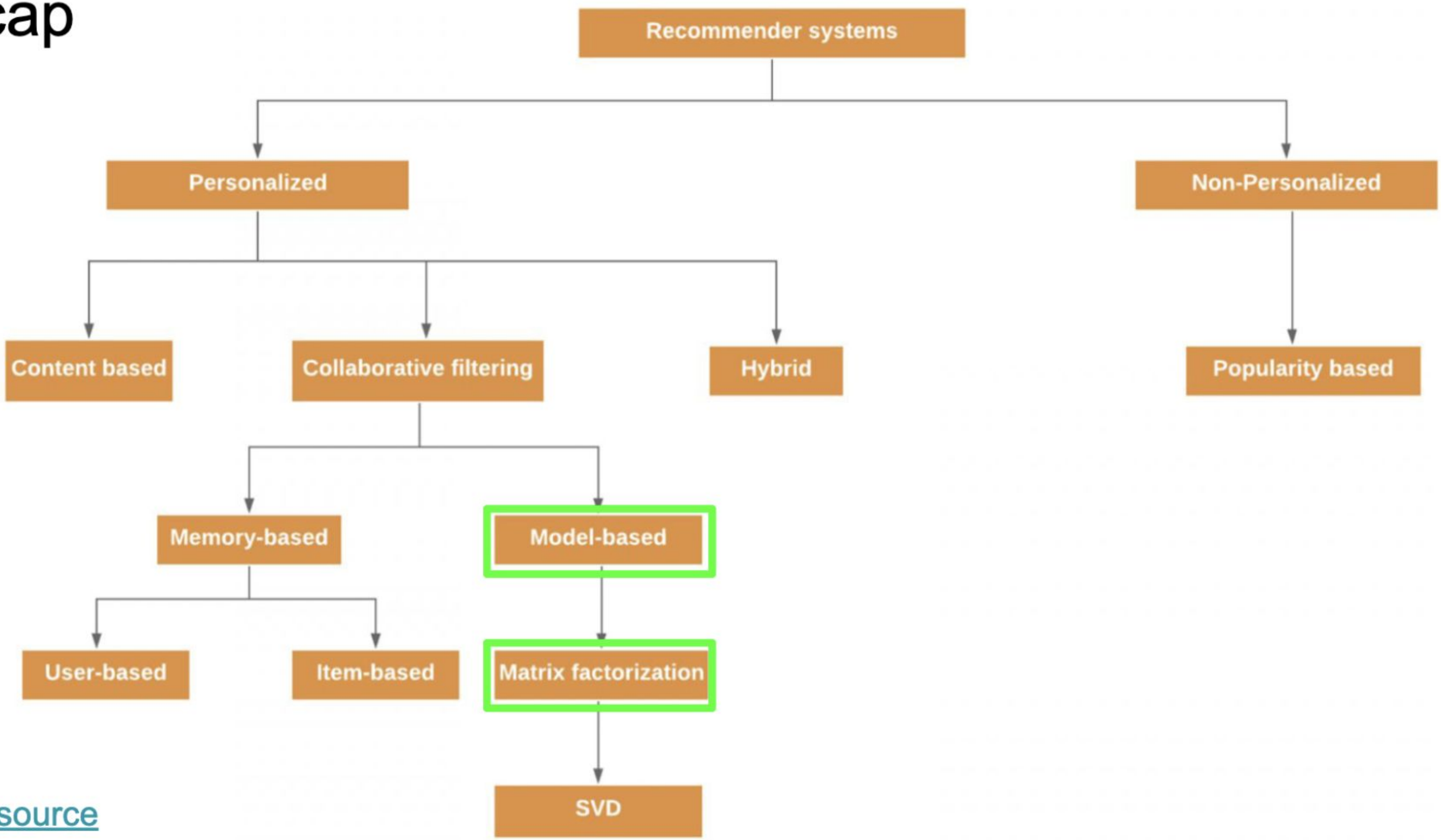


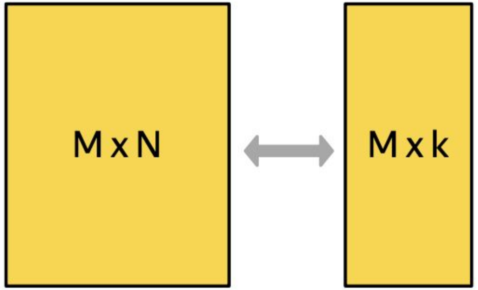
Image source

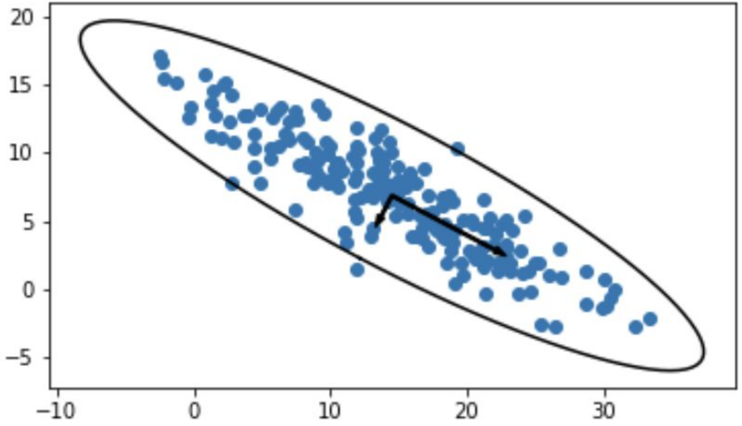
План лекции:

- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- Truncated SVD
- Funk SVD
- Интерпретация SVD
- ALS, IALS

PCA: связь с матричными разложениями

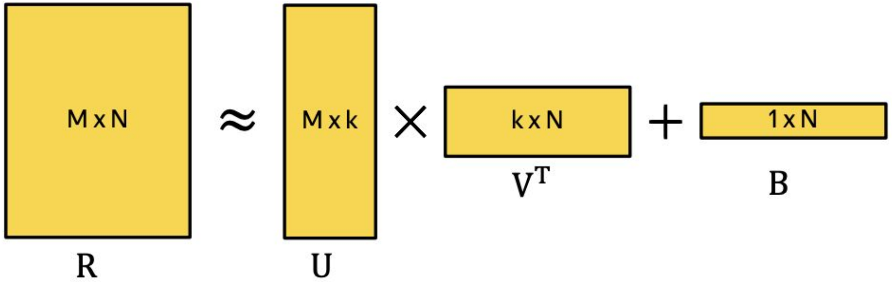
$$\sum_i dist(x_i, L_k) \to \min$$

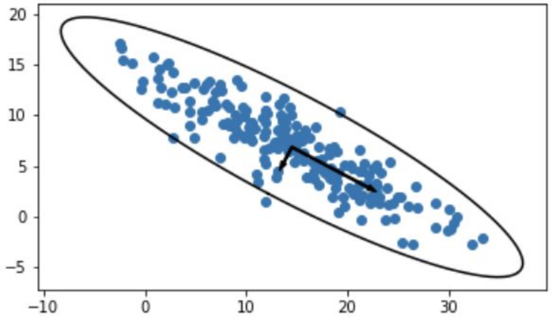




PCA: связь с матричными разложениями

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



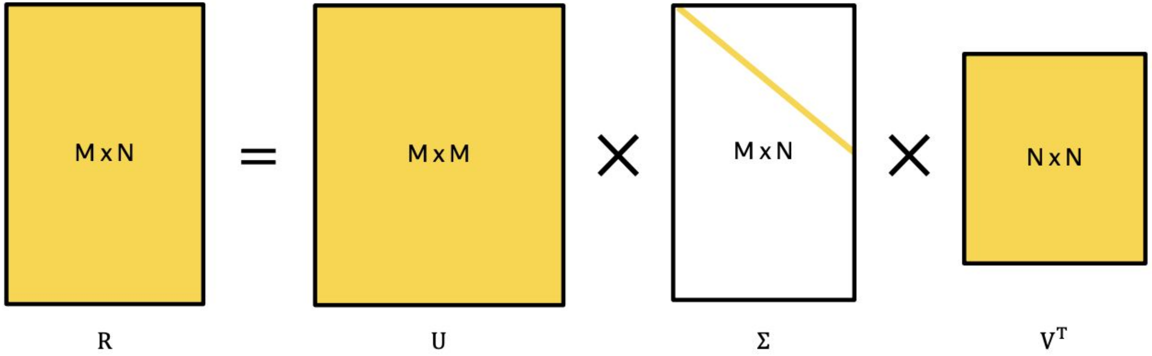


План лекции:

- Recap
- PCA: связь с матричными разложениями
- $\boxed{\text{SVD и SVD compact}}$
- Truncated SVD
- Funk SVD
- Интерпретация SVD
- ALS, IALS

SVD (**Singular value decomposition**)

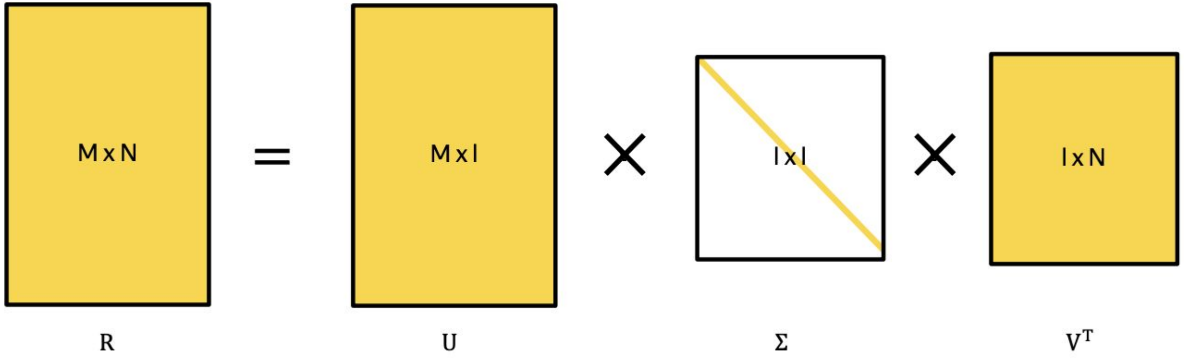
$$R = U\Sigma V^T$$
$$UU^T = U^T U = I_M$$
$$VV^T = V^T V = I_N$$



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

SVD compact

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



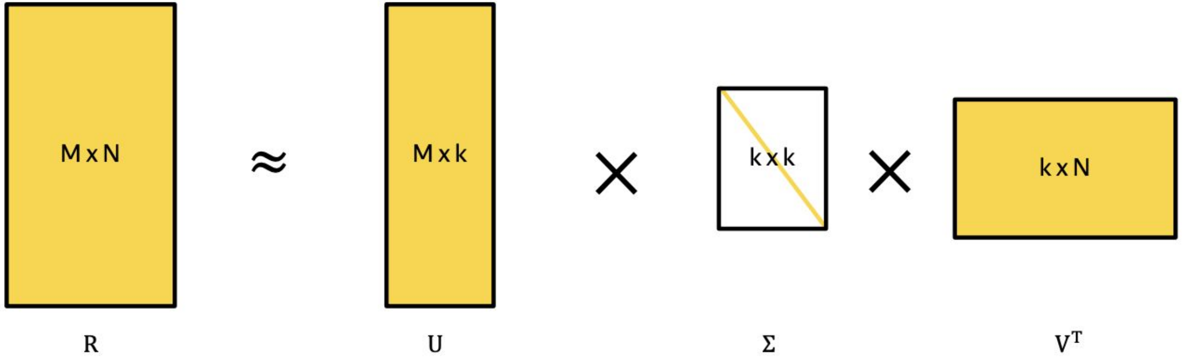
$$r_{ij} = \overline{u_i}^T \Sigma \overline{v_j}$$

План лекции:

- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- $\boxed{\text{Truncated SVD}}$
- Funk SVD
- Интерпретация SVD
- ALS, IALS

Truncated SVD

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



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

Truncated SVD: оптимизация

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

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

Truncated SVD: оптимизация

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

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

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

Truncated SVD: оптимизация

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

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

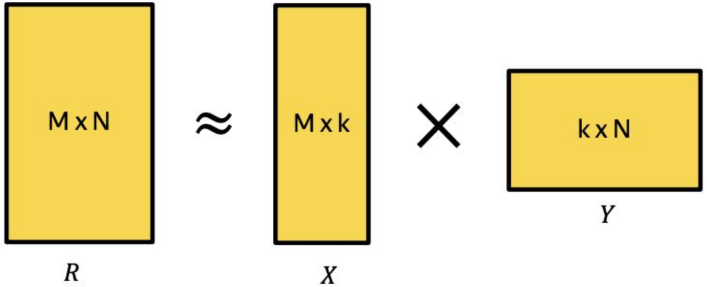
$$\min_{\substack{U,V,\Sigma \\ U^T U = V^T V = I_k \\ \sigma_{ij}=0, i\neq j \\ \sigma_{ii}>\sigma_{jj}, i<j}} \sum_{\forall i,j} \left(r_{ij} - u_i^T \Sigma v_j\right)^2 \quad \Leftrightarrow \quad \min_{X,Y} \sum_{\forall i,j} \left(r_{ij} - x_i^T y_j\right)^2$$

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

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

Truncated SVD: оптимизация

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



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

План лекции:

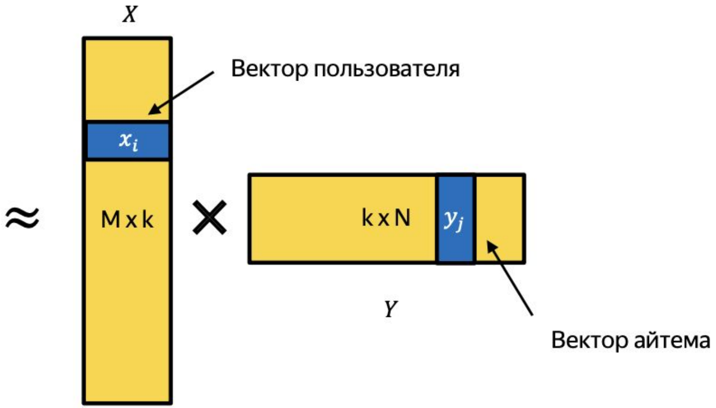
- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- Truncated SVD
- Funk SVD
- Интерпретация SVD
- ALS, IALS

Funk SVD

Айтем

Пользователь

2		2	4	5	
5		4			1
		$\boxed{5}$		2	
	1		5		4
		4			2
4	5		1		



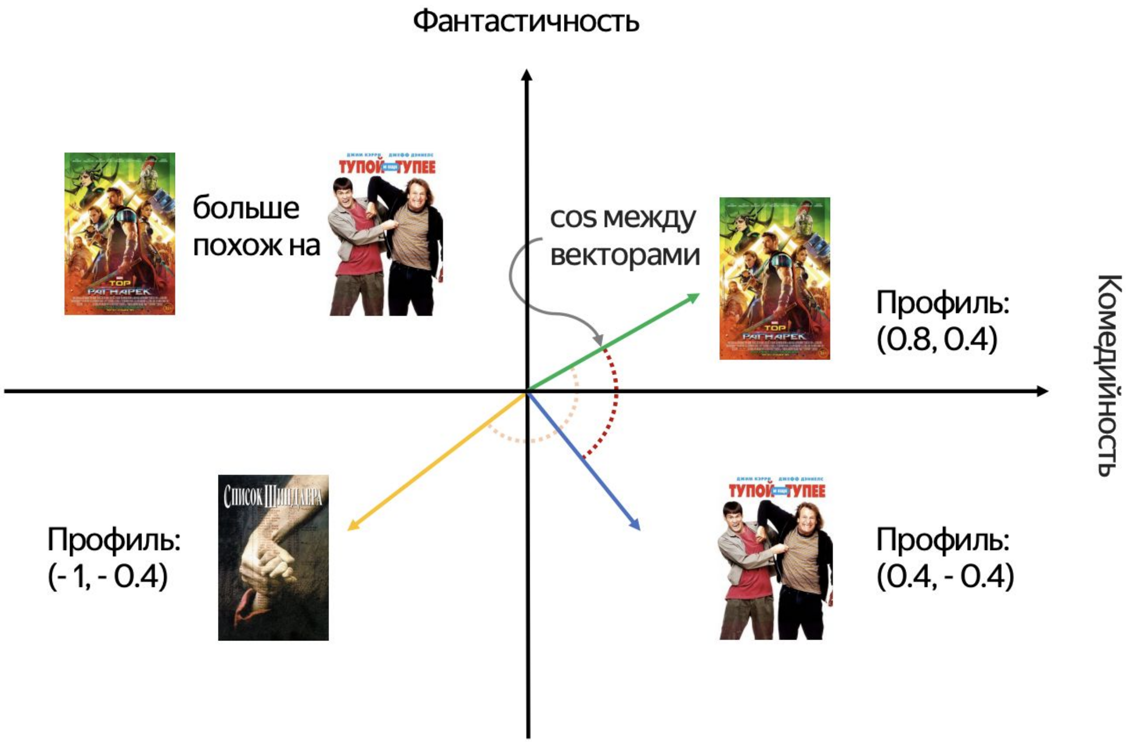
$$\min_{X,Y} \sum_{(i,j)\in R} \left(r_{ij} - x_i^T y_j\right)^2 + \lambda \sum_i \|x_i\|^2 C_i + \lambda \sum_j \left\|y_j\right\|^2 C_j$$

План лекции:

- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- Truncated SVD
- Funk SVD
- $\boxed{\text{Интерпретация SVD}}$
- ALS, IALS

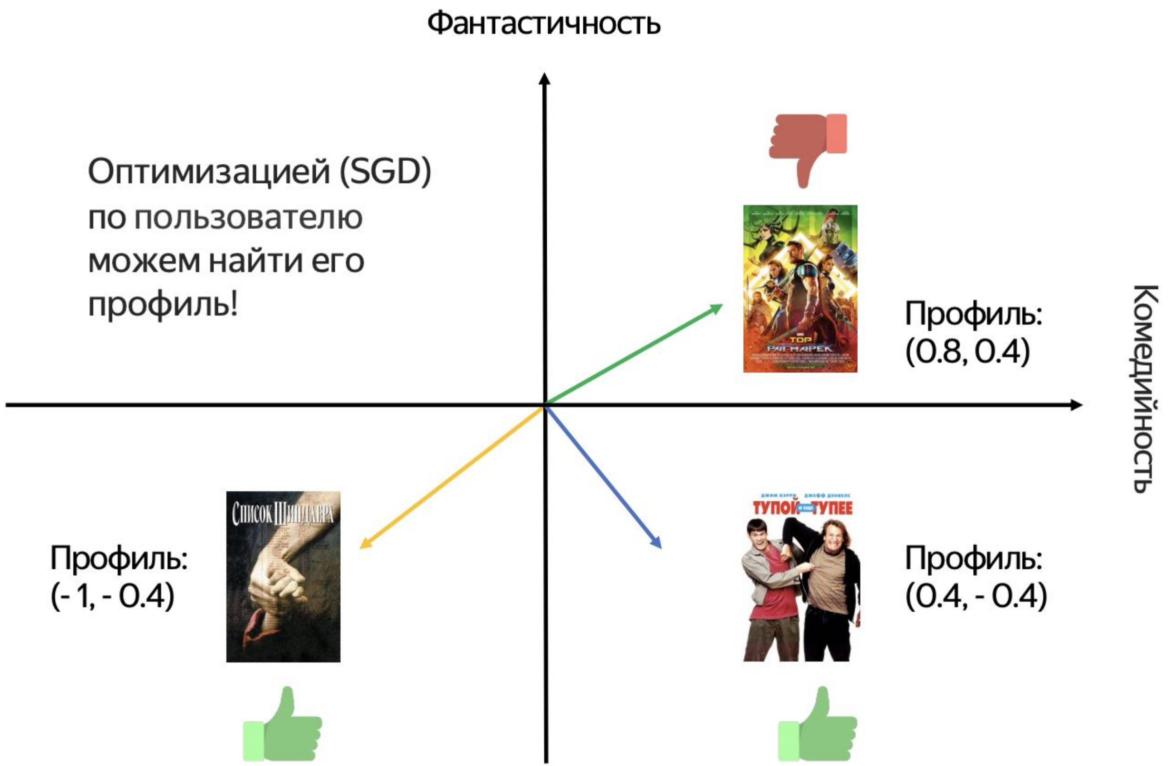
Интерпретация SVD

Интерпретация SVD



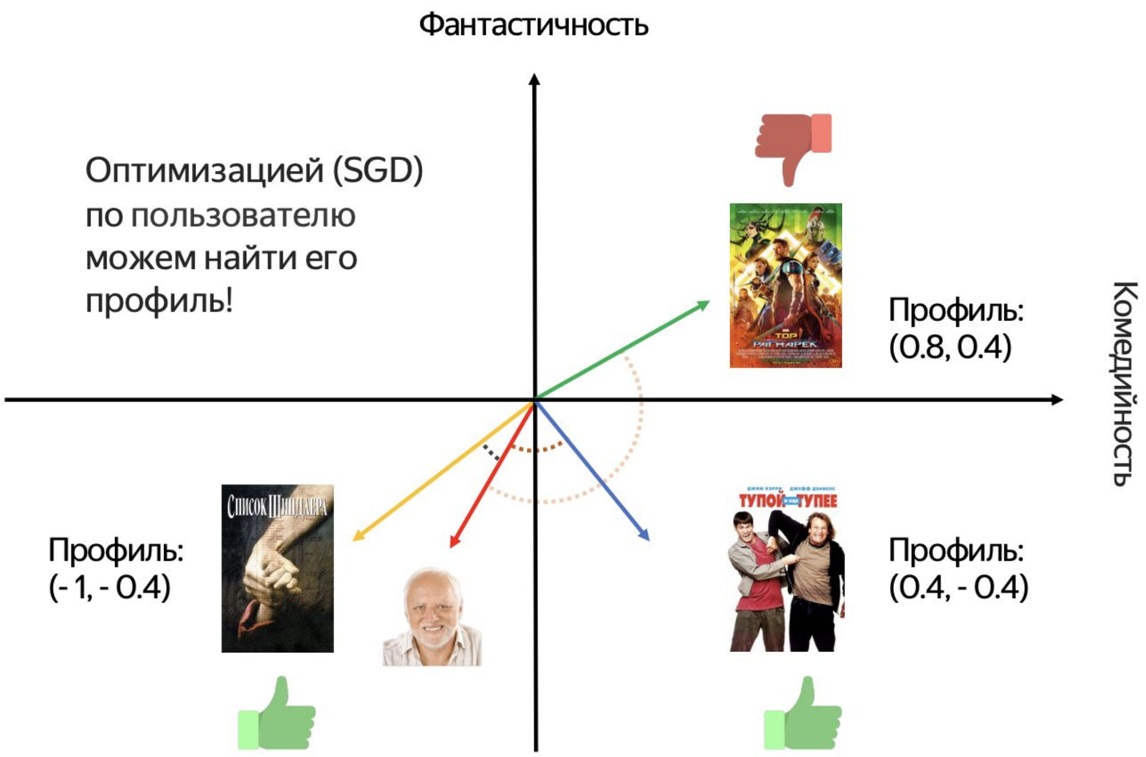
Интерпретация SVD

Оптимизацией (SGD) по пользователю можем найти его профиль!

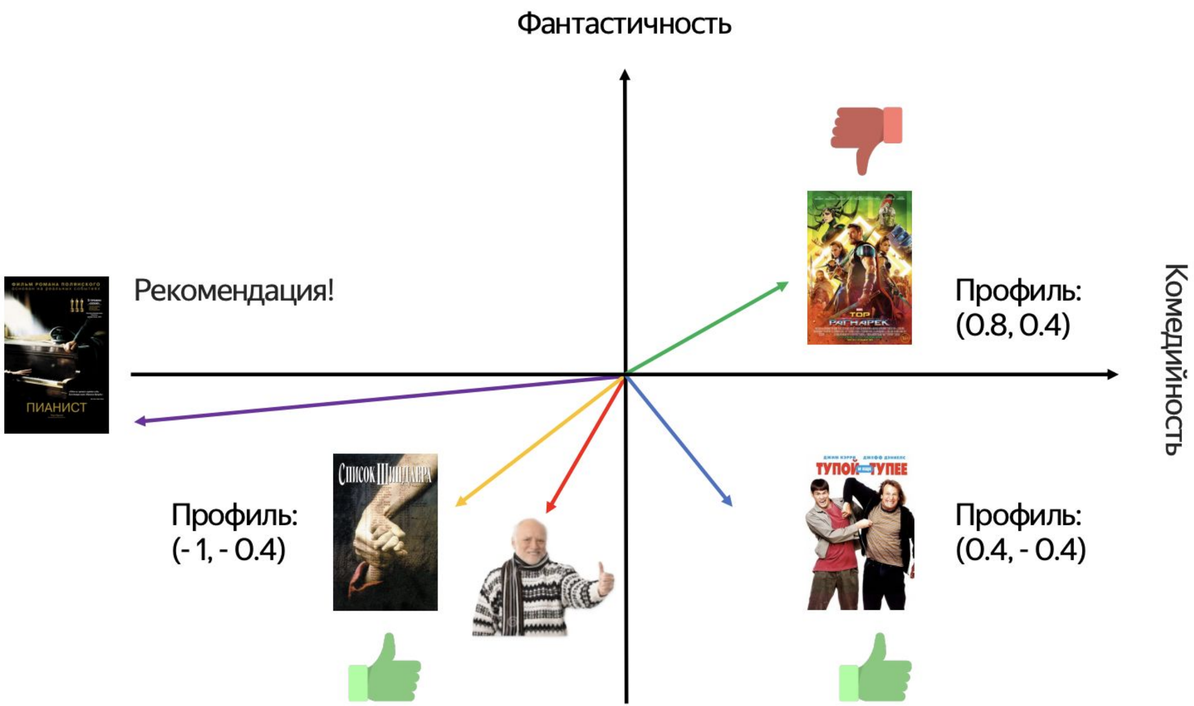


Интерпретация SVD

Оптимизацией (SGD) по пользователю можем найти его профиль!



Интерпретация SVD



Итоги SVD

- В отличие от коллаборативной фильтрации (CF), матричные факторизации (MF) не используют явные сходства между товарами.
- MF стремятся представить пользователей и товары через небольшое число скрытых признаков, объясняющих выставленные оценки.
- Эти признаки часто сложно интерпретировать — как минимум, векторы пользователей и товаров определены лишь с точностью до поворота в пространстве.

План лекции:

- Recap
- PCA: связь с матричными разложениями
- SVD и SVD compact
- Truncated SVD
- Funk SVD
- Интерпретация SVD
- ALS, IALS

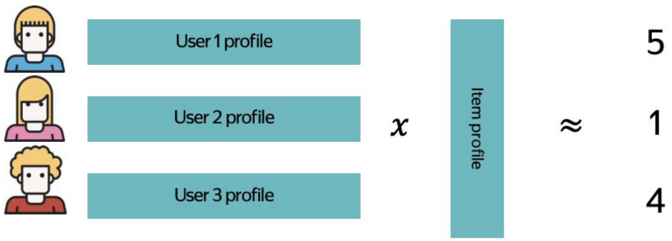
ALS (**Alternating Least Squares**)

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

Инициализируем X и Y случайными значениями.

В цикле:

- Фиксируем матрицу X (пользователей)
- Находим оптимальную матрицу Y (решаем гребневую регрессию для каждого товара)



- И наоборот

ALS (**Alternating Least Squares**)

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

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

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

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

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

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

IALS (**Implicit ALS**)

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

Понравилось ли?

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

Насколько уверены в $\boldsymbol{p_{ij}}$

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

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

IALS (**Implicit ALS**)

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

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

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

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

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