

Data Mining in Action

# Матричные разложения

Виктор Кантор



# Матричные разложения

$$\begin{matrix} X & \approx & U & \cdot & V^T \\ l \times n & & l \times k & & k \times n \end{matrix}$$

# Матричные разложения

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

# Матричные разложения

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$\left\| X - U \cdot V^T \right\| \rightarrow \min$$

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

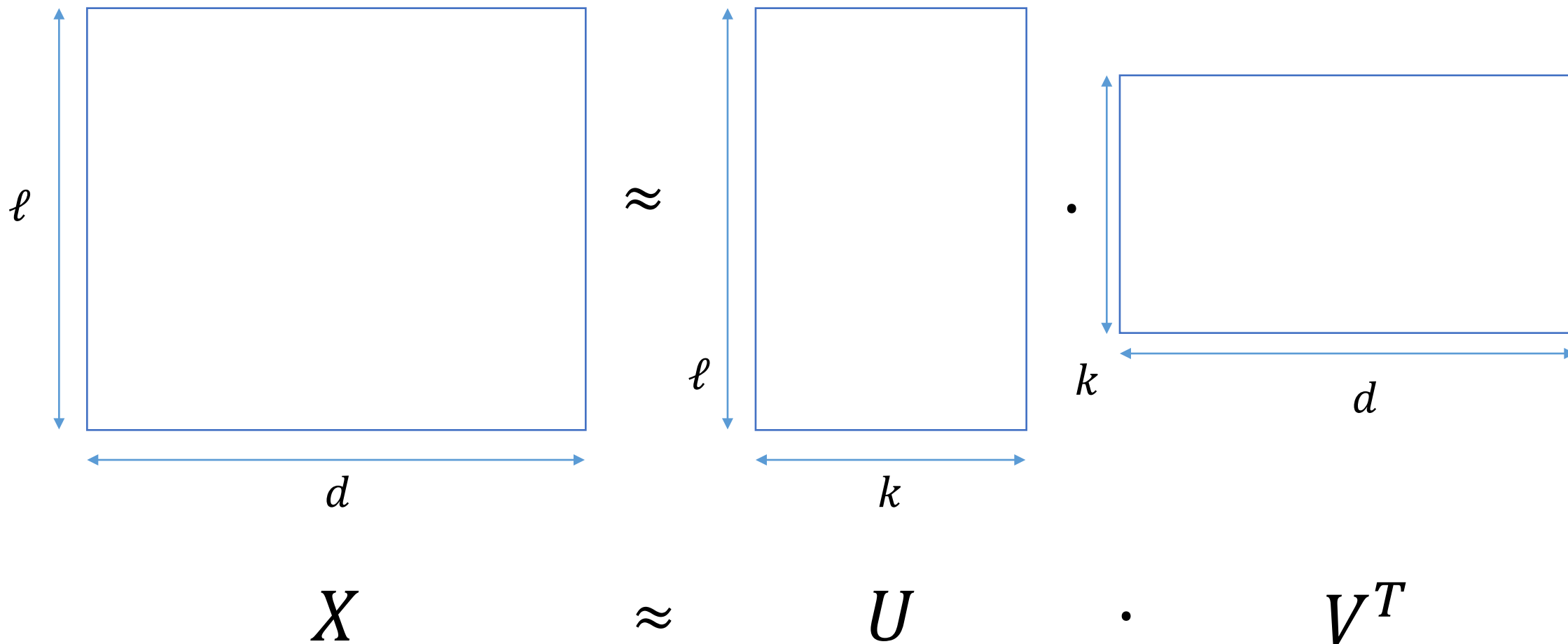
# Матричные разложения

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

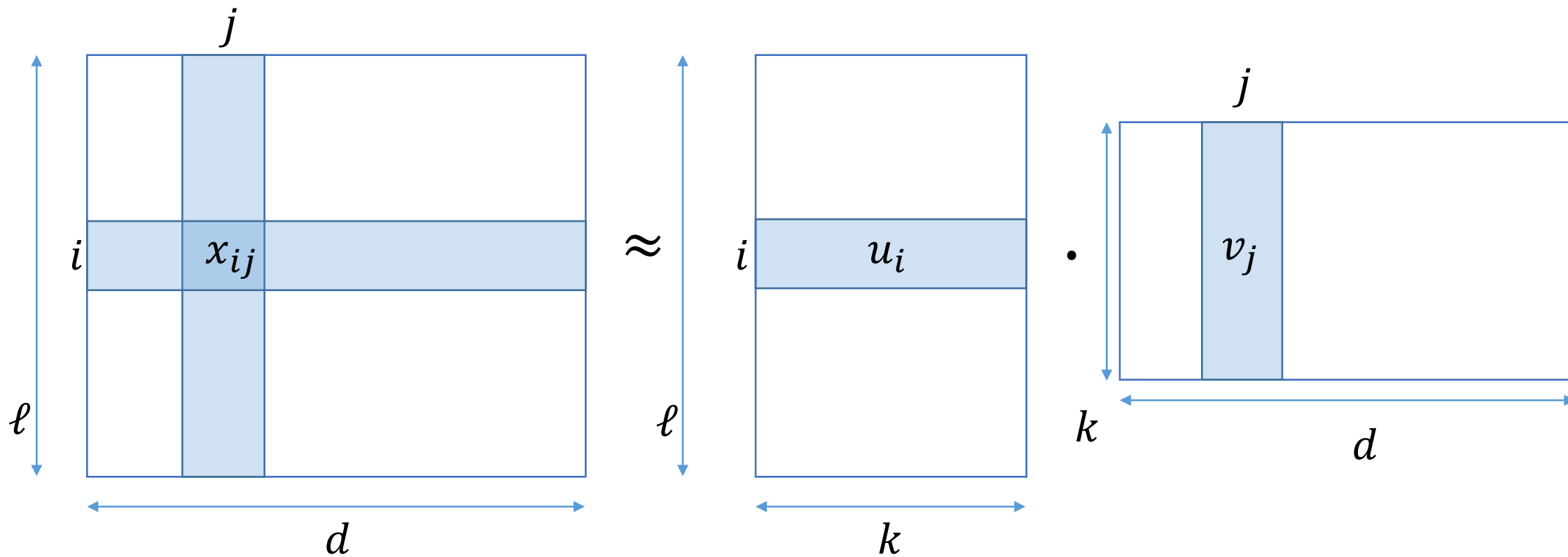
$$||X - U \cdot V^T|| \rightarrow \min$$

$$\sum_{i,j} (x_{ij} - \langle u_i, v_j \rangle)^2 \rightarrow \min$$

# Разбираемся с обозначениями



# Разбираемся с обозначениями



$$x_{ij} \approx \langle u_i, v_j \rangle$$

# SVD в линейной алгебре

$$X = U\Sigma V^T$$

$U$  - ортогональная

$\Sigma$  - диагональная

$V$  - ортогональная



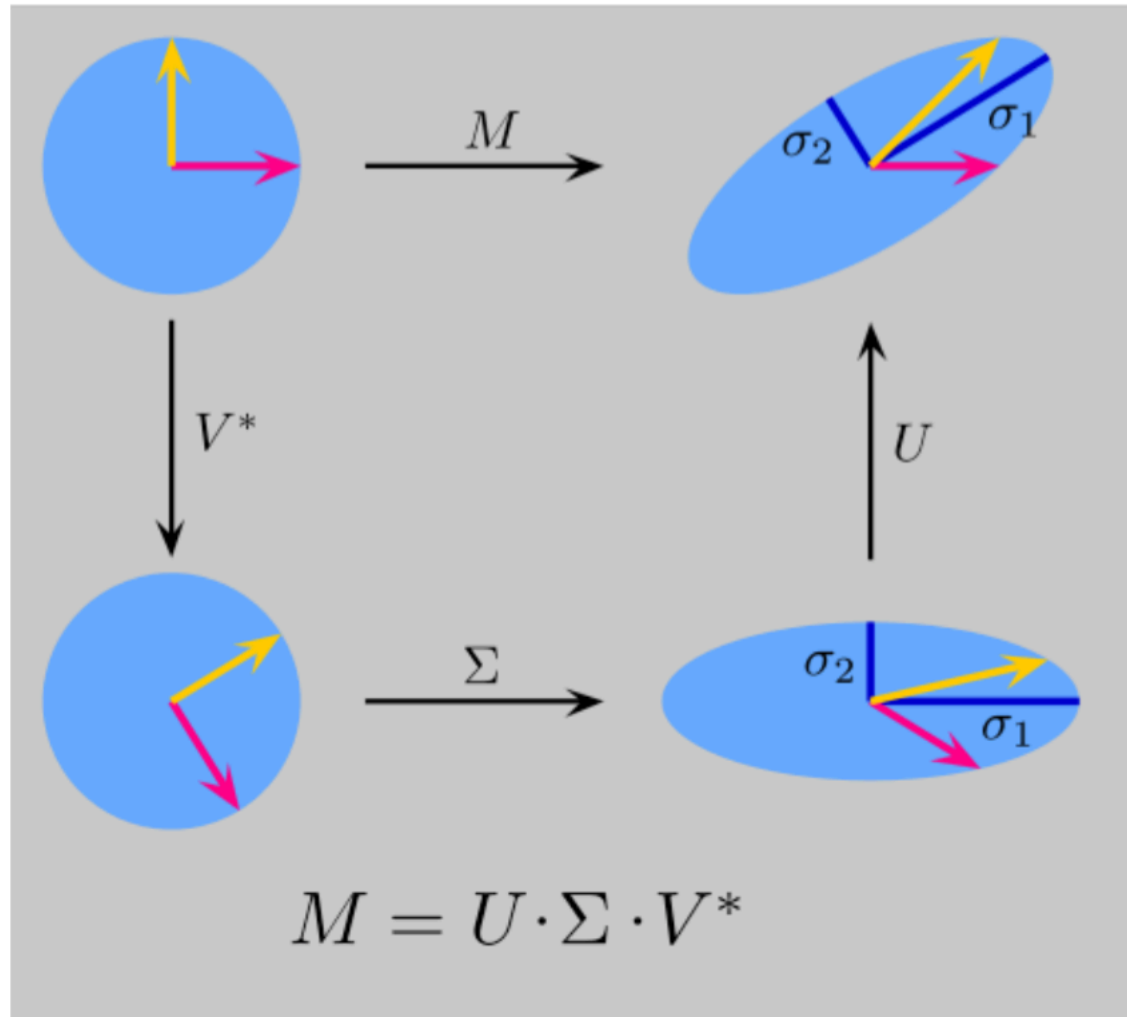
# SVD в линейной алгебре

$$X = U\Sigma V^T$$

$U$  - ортогональная

$\Sigma$  - диагональная

$V$  - ортогональная



# Приближение матрицы с помощью SVD

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

Приближение матрицы с помощью SVD

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

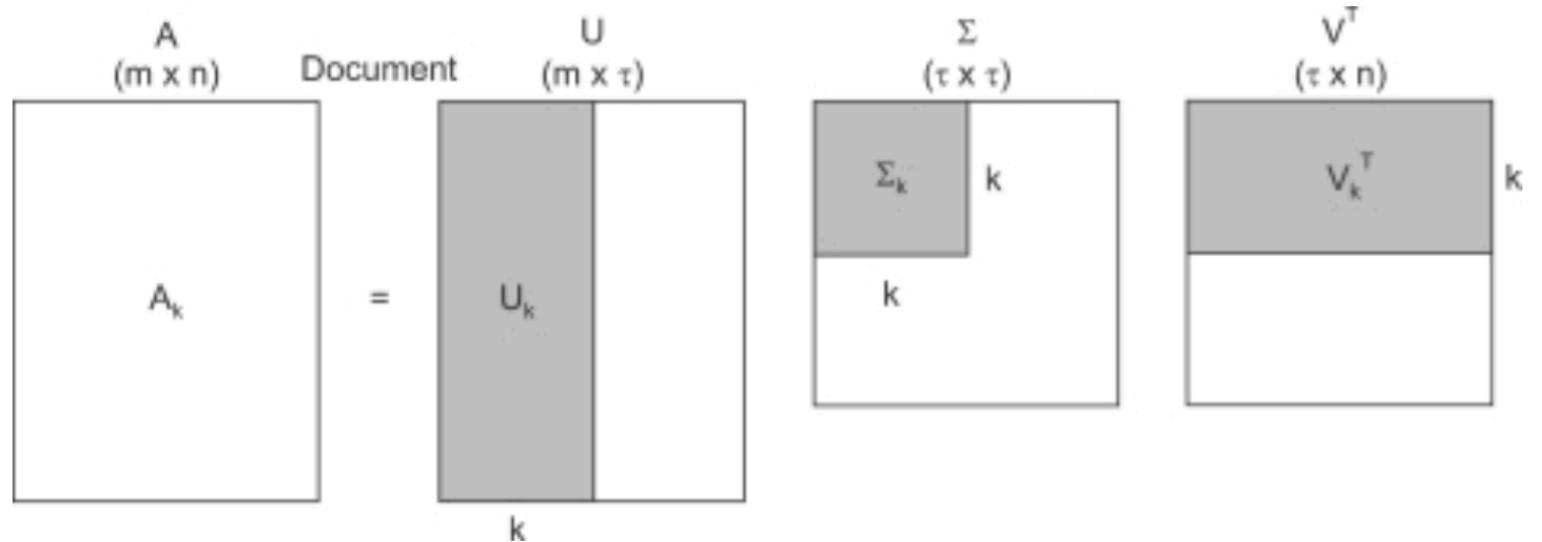
$$X = \tilde{U} \Sigma \tilde{V}^T$$

# Приближение матрицы с помощью SVD

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

$$X = \tilde{U} \Sigma \tilde{V}^T$$



# Приближение матрицы с помощью SVD

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

$$X = \tilde{U} \Sigma \tilde{V}^T$$

$\tilde{U}_k, \Sigma_k, \tilde{V}_k$ -усеченные матрицы из SVD

$$U = \tilde{U}_k \Sigma_k, \quad V = \tilde{V}_k$$

# Приближение матрицы с помощью SVD

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

$$X = \tilde{U} \Sigma \tilde{V}^T$$

$\tilde{U}_k, \Sigma_k, \tilde{V}_k$ -усеченные матрицы из SVD

$$U = \tilde{U}_k, \quad V = \tilde{V}_k \Sigma_k$$

# Приближение матрицы с помощью SVD

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$||X - U \cdot V^T|| \rightarrow \min$$

$$X = \tilde{U} \Sigma \tilde{V}^T$$

$\tilde{U}_k, \Sigma_k, \tilde{V}_k$ -усеченные матрицы из SVD

$$U = \tilde{U}_k \sqrt{\Sigma_k}, \quad V = \tilde{V}_k \sqrt{\Sigma_k}$$

# ”SVD” в машинном обучении

$$\underset{l \times n}{X} \approx \underset{l \times k}{U} \cdot \underset{k \times n}{V^T}$$

$$\sum_{i,j} (x_{ij} - \langle u_i, v_j \rangle)^2 \rightarrow \min$$

$u_i$  - «профили» объектов

$v_j$  - «профили» исходных признаков



# Матрица рейтингов и SVD

	Пила	Улица Вязов	Ванильное небо	1+1
Маша	5	4	1	2
Юля	5	5	2	
Вова			3	5
Коля	3		4	5
Петя				4
Ваня		5	3	3

# Матрица рейтингов и SVD

		$j$			
$i$		Пила	Улица Вязов	Ванильное небо	1+1
	Маша	5	4	1	2
	Юля	5	5	2	
	Вова			3	5
	Коля	3		4	5
	Петя				4
	Ваня		5	3	3

# Матрица рейтингов и SVD

		<i>j</i>			
		Пила	Улица Вязов	Ванильное небо	1+1
<i>i</i>	Маша	5	4	1	2
	Юля	5	5	2	
	Вова			3	5
	Коля	3		4	5
	Петя				4
	Ваня		5	3	3

$$x_{ij} \approx \langle u_i, v_j \rangle$$

$u_i$  - «интересы пользователей»

$v_j$  - «параметры фильмов»

# Матрица частот слов и SVD

	database	SQL	index	regression	likelihood	linear
d1	24	21	9	0	0	3
d2	32	10	5	0	3	0
d3	12	16	5	0	0	0
d4	6	7	2	0	0	0
d5	43	31	20	0	3	0
d6	2	0	0	18	7	16
d7	0	0	1	32	12	0
d8	3	0	0	22	4	2
d9	1	0	0	34	27	25
d10	6	0	0	17	4	23

# Матрица частот слов и SVD $j$

$i$

	database	SQL	index	regression	likelihood	linear
d1	24	21	9	0	0	3
d2	32	10	5	0	3	0
d3	12	16	5	0	0	0
d4	6	7	2	0	0	0
d5	43	31	20	0	3	0
d6	2	0	0	18	7	16
d7	0	0	1	32	12	0
d8	3	0	0	22	4	2
d9	1	0	0	34	27	25
d10	6	0	0	17	4	23

# Матрица частот слов и SVD $j$

	database	SQL	index	regression	likelihood	linear
d1	24	21	9	0	0	3
d2	32	10	5	0	3	0
d3	12	16	5	0	0	0
d4	6	7	2	0	0	0
$i$ d5	43	31	20	0	3	0
d6	2	0	0	18	7	16
d7	0	0	1	32	12	0
d8	3	0	0	22	4	2
d9	1	0	0	34	27	25
d10	6	0	0	17	4	23

$$x_{ij} \approx \langle u_i, v_j \rangle$$

$u_i$  - «темы» документов

$v_j$  - «темы» слов

# Постановка задачи

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min_{u_i, v_j}$$

# Градиентный спуск (GD)

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min_{u_i, v_j}$$

$$\frac{\partial Q}{\partial u_i} = \sum_{\tilde{i}, j} \frac{\partial}{\partial u_i} (\langle u_{\tilde{i}}, v_j \rangle - x_{\tilde{i}j})^2 = \sum_j 2(\langle u_i, v_j \rangle - x_{ij}) \frac{\partial \langle u_i, v_j \rangle}{\partial u_i} =$$

$$= \sum_j 2(\langle u_i, v_j \rangle - x_{ij}) v_j \quad \varepsilon_{ij} = (\langle u_i, v_j \rangle - x_{ij}) - \text{ошибка на } x_{ij}$$

$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \sum_j \varepsilon_{ij} v_j$$

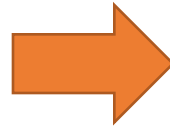


# Стохастический градиентный спуск (SGD)

GD:

$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \sum_j \varepsilon_{ij} v_j$$

$$v_j^{(t+1)} = v_j^{(t)} - \eta_t \sum_i \varepsilon_{ij} u_i$$



SGD:

$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \varepsilon_{ij} v_j$$

$$v_j^{(t+1)} = v_j^{(t)} - \eta_t \varepsilon_{ij} u_i$$

Для случайных  $i, j$

# Плюсы и минусы SGD

+Простота реализации

+Сходимость

- Медленно сходится

- Сложность выбора шага градиентного спуска ( $\gamma_t$  и  $\eta_t$ )

- При константном шаге сходится очень медленно

# Идея ALS

$$Q \rightarrow \min_{u_i, v_j}$$

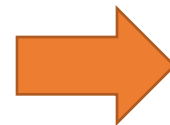
Повторяем до сходимости:

$$\frac{\partial Q}{\partial u_i} = 0$$



$u_i$

$$\frac{\partial Q}{\partial v_j} = 0$$



$v_j$

Выписываем шаг в ALS

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min_{u_i, v_j}$$

$$\frac{\partial Q}{\partial u_i} = \sum_j 2(\langle u_i, v_j \rangle - x_{ij})v_j = 0 \quad \sum_j v_j \langle v_j, u_i \rangle = \sum_j x_{ij}v_j$$

$$\sum_j v_j v_j^T u_i = \sum_j x_{ij} v_j$$
$$\underbrace{\left( \sum_j v_j v_j^T \right)}_A u_i = \underbrace{\sum_j x_{ij} v_j}_b$$

# ALS: итоговый алгоритм

Повторяем по случайным  $i, j$  до сходимости:

$$\left( \sum_j v_j v_j^T \right) u_i = \sum_j x_{ij} v_j \quad \Rightarrow \quad u_i \quad \text{(решение системы линейных уравнений)}$$

$$\left( \sum_i u_i u_i^T \right) v_j = \sum_i x_{ij} u_i \quad \Rightarrow \quad v_j$$

# Регуляризация

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 \rightarrow \min_{u_i, v_j}$$

$\alpha$  и  $\beta$  - небольшие положительные числа (0.001, 0.01, 0.05)

Оптимизируемый функционал

$$x_{ij} \approx \langle u_i, v_j \rangle$$

$$\sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min$$

# Сдвиг

$$x_{ij} \approx \boxed{\mu} + \langle u_i, v_j \rangle$$

$$\sum_{i,j} (\boxed{\mu} + \langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min$$



# Базовые предикторы

$$x_{ij} \approx \mu + \boxed{b_i^u} + \boxed{b_j^v} + \langle u_i, v_j \rangle$$

$$\sum_{i,j} \left( \mu + \boxed{b_i^u} + \boxed{b_j^v} + \langle u_i, v_j \rangle - x_{ij} \right)^2 \rightarrow \min$$

# Регуляризация

$$\sum_{i,j} (\mu + b_i^u + b_j^v + \langle u_i, v_j \rangle - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 +$$

$+ \gamma \sum_i b_i^{u^2} + \delta \sum_j b_j^{v^2} \rightarrow \min$

# И снова рекомендации: implicit разложения

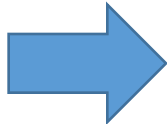
$j$

	Вечернее платье	Поднос для писем	iPhone 6s	Шуба D&G
	Маша	1		1
	Юля	1		1
	Вова	1	1	
$i$	Коля	1	1	
	Петя	1	1	
	Ваня		1	1

# Почему нужно что-то менять

		$j$			
		Вечернее платье	Поднос для писем	iPhone 6s	Шуба D&G
$i$	Маша	1		1	
	Юля	1	1		1
	Вова		1	1	
	Коля	1	?	1	
	Петя		1	1	
	Ваня			1	1

$$x_{ij} = 1 \approx \langle u_i, v_j \rangle$$

$$\sum_{i,j: x_{ij} \neq 0} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min$$


$$u_i = \frac{1}{\sqrt{d}} (1 \quad \dots \quad 1)$$

$$v_j = \frac{1}{\sqrt{d}} (1 \quad \dots \quad 1)$$

# Explicit и implicit

- **Explicit feedback:** есть положительные и отрицательные пример (например, низкие и высокие оценки фильмов, лайки и дислайки и т.д.)
- **Implicit feedback:** есть только положительные (покупки, просмотры, лайки) или только отрицательные примеры (дислайки)

# Implicit matrix factorization

$$\sum_{i,j} w_{ij} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min$$



Сумма по всем индексам (не только по известным элементам матрицы)

$w_{ij}$  принимает большие значения для  $x_{ij} \neq 0$   
и значительно меньшие для  $x_{ij} = 0$

# Implicit ALS

$$\sum_{i,j} w_{ij} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min$$

$$w_{ij} = 1 + \alpha |x_{ij}| \quad \alpha = 10, 100, 1000$$

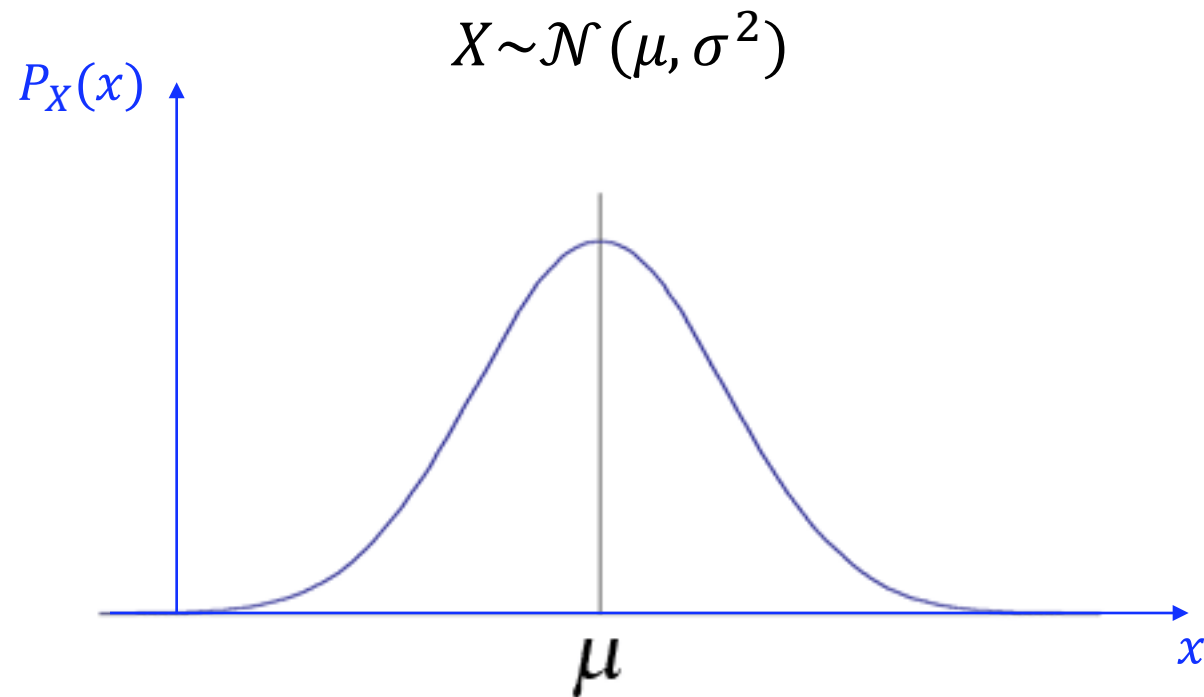
$u_i, v_j$  оцениваем с помощью ALS

# Постановка задачи в "SVD"

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min_{u_i, v_j}$$



# Нормальное распределение



$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

# Связь "SVD" и нормального распределения

$$x_{ij} = \langle u_i, v_j \rangle + \varepsilon \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$x_{ij} \sim \mathcal{N}(\langle u_i, v_j \rangle, \sigma^2)$$

$$\prod_{i,j} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_{ij} - \langle u_i, v_j \rangle)^2}{2\sigma^2}} \rightarrow \max$$

$$\sum_{i,j} \frac{(x_{ij} - \langle u_i, v_j \rangle)^2}{2\sigma^2} - \frac{1}{2} \ln 2\pi\sigma^2 \rightarrow \min$$

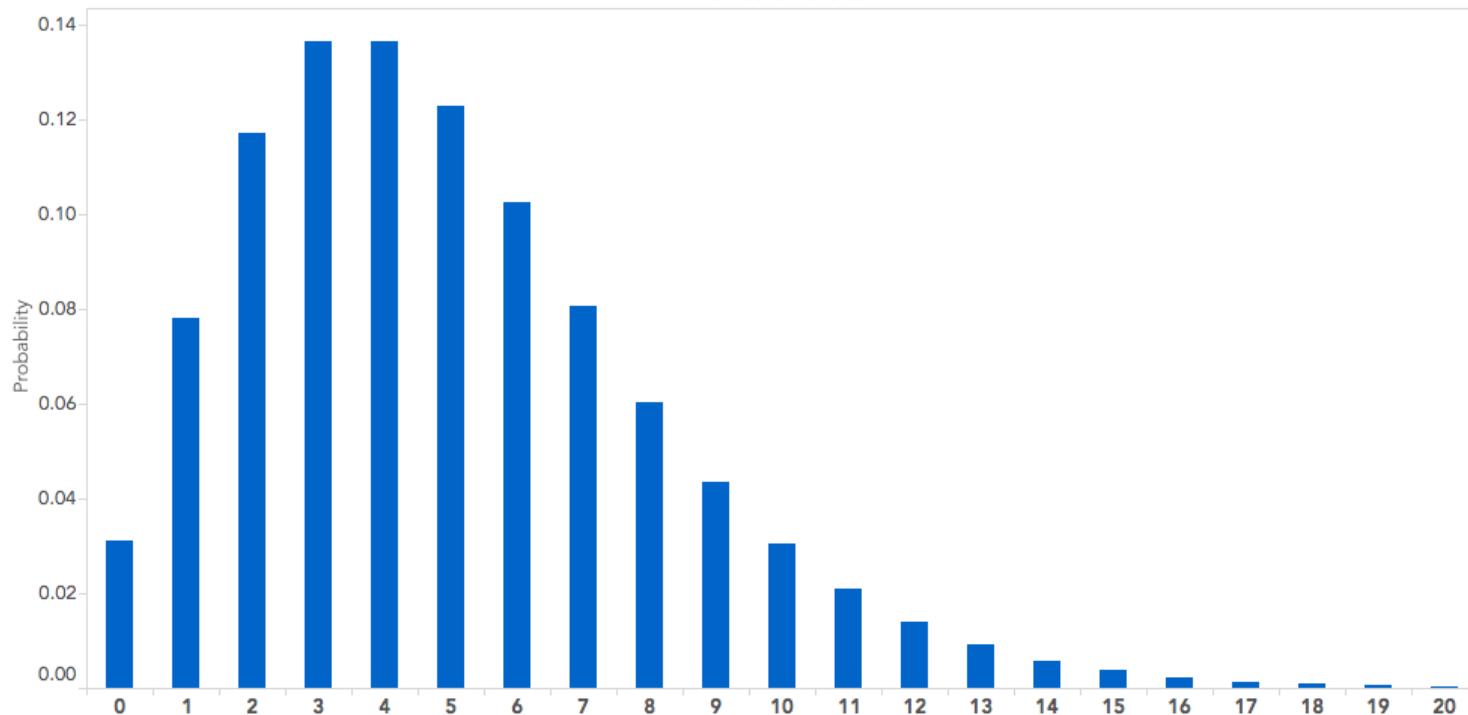
$$\sum_{i,j} (x_{ij} - \langle u_i, v_j \rangle)^2 \rightarrow \min$$

# Какое распределение подходит больше

	database	SQL	index	regression	likelihood	linear
d1	24	21	9	0	0	3
d2	32	10	5	0	3	0
d3	12	16	5	0	0	0
d4	6	7	2	0	0	0
d5	43	31	20	0	3	0
d6	2	0	0	18	7	16
d7	0	0	1	32	12	0
d8	3	0	0	22	4	2
d9	1	0	0	34	27	25
d10	6	0	0	17	4	23

# Распределение Пуассона

$$X \sim \text{Pois}(\lambda)$$



$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \mathbb{E}X = \lambda$$

# Распределение Пуассона и матричные разложения

$$x_{ij} \sim \text{Pois}(\langle u_i, v_j \rangle) \quad P(x_{ij}) = \frac{\langle u_i, v_j \rangle^{x_{ij}}}{x_{ij}!} e^{-\langle u_i, v_j \rangle}$$

$$\prod_{i,j} \frac{\langle u_i, v_j \rangle^{x_{ij}}}{x_{ij}!} e^{-\langle u_i, v_j \rangle} \rightarrow \max$$

$$\sum_{i,j} \langle u_i, v_j \rangle - x_{ij} \ln \langle u_i, v_j \rangle + \ln x_{ij}! \rightarrow \min$$

$$\sum_{i,j} \langle u_i, v_j \rangle - x_{ij} \ln \langle u_i, v_j \rangle \rightarrow \min$$

# SGD для NMF (Non-negative matrix factorization)

$$Q = \sum_{i,j} \langle u_i, v_j \rangle - x_{ij} \ln \langle u_i, v_j \rangle \rightarrow \min$$

$$\frac{\partial Q}{\partial u_i} = \sum_j v_j - \frac{x_{ij}}{\langle u_i, v_j \rangle} v_j = \sum_j \underbrace{\frac{\langle u_i, v_j \rangle - x_{ij}}{\langle u_i, v_j \rangle}}_{\tilde{\varepsilon}_{ij} - \text{«относительная ошибка» прогноза}} v_j \rightarrow \min$$

SGD:

$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \tilde{\varepsilon}_{ij} v_j$$

$$v_j^{(t+1)} = v_j^{(t)} - \eta_t \tilde{\varepsilon}_{ij} u_i$$

# Другие неотрицательные матричные разложения

Можно использовать норму Фробениуса, но добавить ограничения неотрицательности для  $U$  и  $V$ :

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \rightarrow \min_{\substack{u_i, v_j: \\ u_{ik} \geq 0 \\ v_{jk} \geq 0}}$$