

Машинное обучение

Лекция 10

Рекомендательные системы

Михаил Гуцин

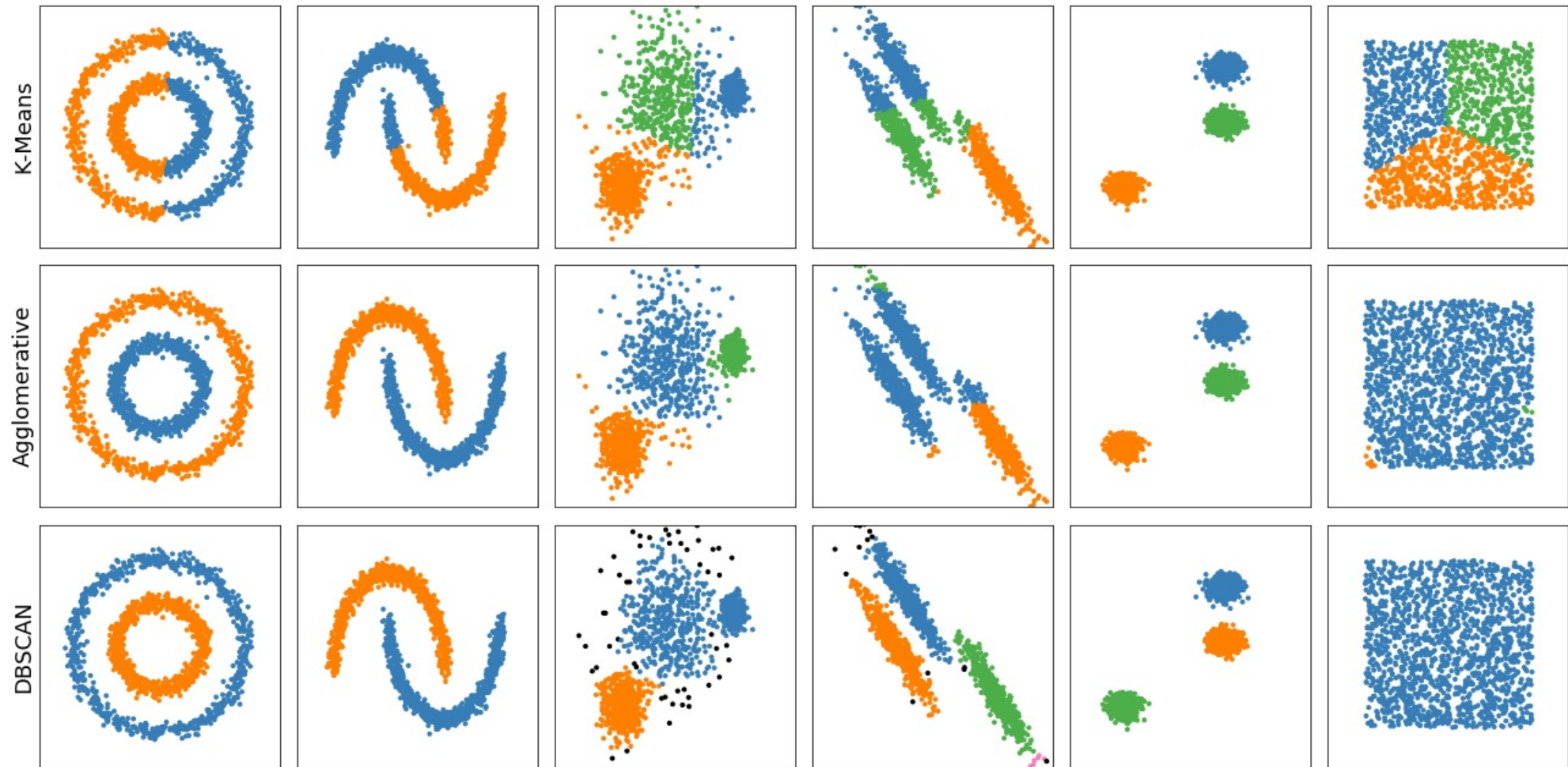
mhushchyn@hse.ru

НИУ ВШЭ, 2022



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

На прошлой лекции



План


- ▶ Коллаборативная фильтрация (memory-based)
 - User-based filtering
 - Item-based filtering
- ▶ Модели со скрытыми переменными
 - Latent factor model
- ▶ Factorization Machines

Введение




Рекомендации фильмов


Рекомендации >




Шантарам
2022, боевик




Китайский квартал
1974, триллер




Штамм
2014, ужасы



Отпетые мошенницы
2019, комедия



Наша Маша и Волшебный орех
2009, мультфильм



Глаз
2008, ужасы

Источник: <https://www.kinopoisk.ru>

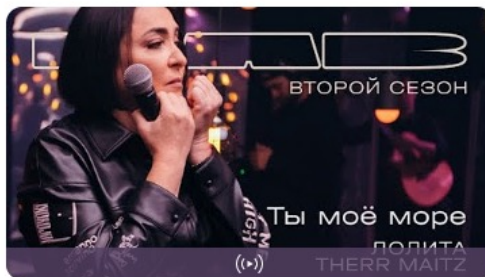
Рекомендации музыки

Все Джемьы Музыка Сейчас в эфире Скetch-шоу Видеоигры Математика Природа Изобразительное искусство Экшен и приключения Футбол Кулинария Последние опубликовано >



Джем – Елизавета Пурис. «Is That Alright?» - Слепые прослушивания - Голос - Сезон 9

Елизавета Пурис, Александра Болдарева, Яна Габбасова и другие исполнители



Джем – Поп-музыка

Therr Maitz, Лолита Марковна Милявская, Антон Беляев и другие исполнители



KRISTINA SI - Мой мармеладный (кавер на Катю Лель) | Битва...

МУЗ-ТВ
239 тыс. просмотров • 2 недели назад



Zivert - Vinyl#2 | Album | 2021

Первое Музыкальное
2,4 млн просмотров • 1 год назад



DS CREW | PROFI | MOVE FORWARD 2022

MOVE FORWARD
285 тыс. просмотров • 5 месяцев назад



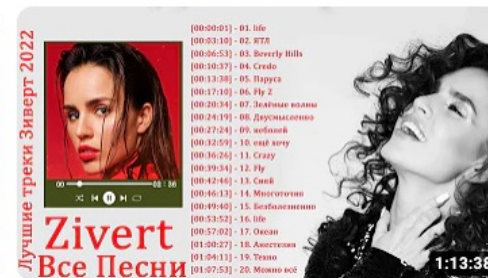
Ваня Дмитриенко - 31-я весна | БИТВА ПОКОЛЕНИЙ

МУЗ-ТВ
568 тыс. просмотров • 1 месяц назад



Zivert | VK Fest 2022 в Санкт Петербурге

ВКонтакте
274 тыс. просмотров • 2 месяца назад



Zivert Лучшие треки 2022
полный альбом

Russian Music
292 тыс. просмотров • 7 месяцев назад

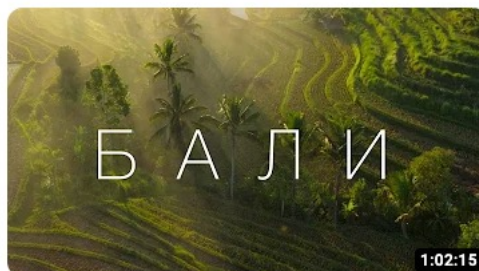
Источник: <https://www.youtube.com>

И даже рекомендации лекций

Все Джеммы Музыка Сейчас в эфире Скetch-шоу Видеоигры Математика **Природа** Изобразительное искусство Экшен и приключения Футбол Кулинария Последние опубликованы >



Путешествие к началу времён
KOSMO ✓
10 млн просмотров • 8 месяцев назад



Бали и почему все на нём помешались. Большой Выпуск.
Антон Птушкин ✓
32 млн просмотров • 2 года назад



Что скрывает ВЕЧНАЯ МЕРЗЛОТА?
Зомби, вирусы и дыры в земле /...
Редакция.Наука ✓
318 тыс. просмотров • 3 месяца назад



What Really Is Everything?
History of the Universe ✓
2,7 млн просмотров • 1 год назад



Fly Into Your Pictures With AI!
InfiniteNature-Zero
What's AI by Louis Bouchard
1,2 тыс. просмотров • 5 дней назад



Путешествие на край Вселенной
Meditat1
1 млн просмотров • 3 года назад



Лекция Сергея Попова "Математика и законы природы" в рамках цикла...
Центр Архэ
84 тыс. просмотров
• Трансляция закончилась 7 месяцев



Старинец А. От Теории Струн к Нейтронным Звёздам и Чёрным...
ПолитСигнал!
15 тыс. просмотров • 3 недели назад

Источник: <https://www.youtube.com>

Мотивация

Каждый день мы принимаем множество решений:























- ▶ Какую музыку слушать?
- ▶ Какой фильм смотреть?
- ▶ Какую книгу читать?
- ▶ Какую еду есть?
- ▶ Какую одежду надеть?
- ▶ В какую игру играть?
- ▶ На какую лекцию идти? 😊



Коллаборативная фильтрация



Задача

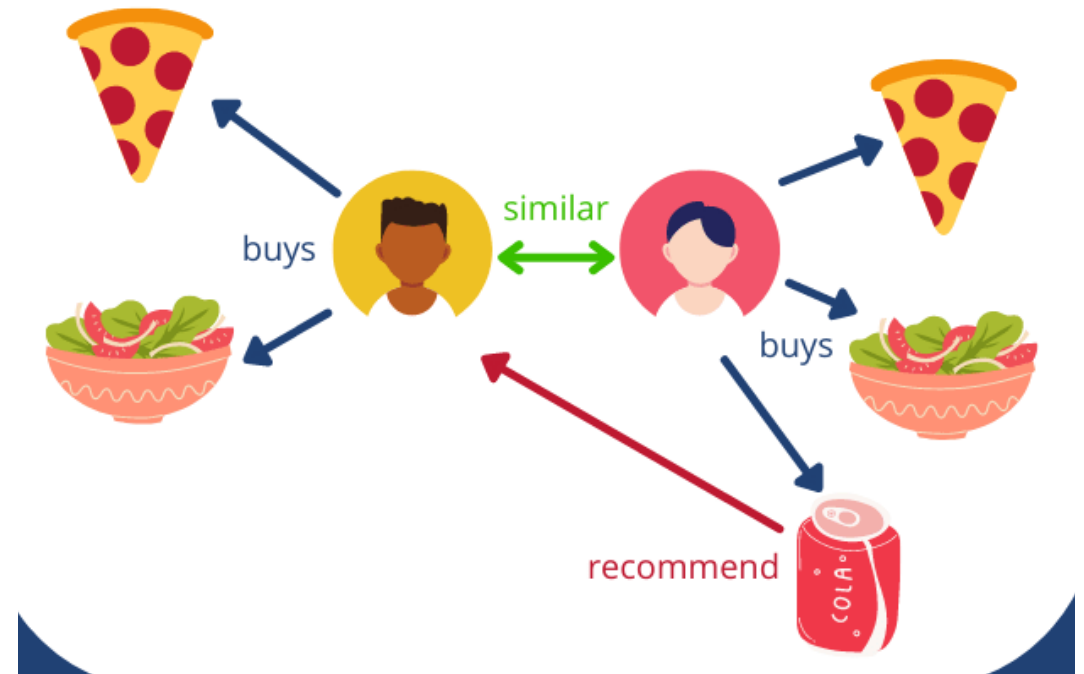
		 Book 1	 Book 2	 Book 3	 Book 4	 Book 5
 User A						
 User B						
 User C						
 User D						

Обозначения























- ▶ U – множество всех пользователей (**U**users)
- ▶ I – множество всех товаров (**I**tems)
- ▶ u, v – некоторые пользователи
- ▶ i, j – некоторые товары
- ▶ U_i – все пользователи, которые оценили товар i
- ▶ I_u – все товары, которые оценил пользователь u
- ▶ R_{ui} – истинный рейтинг (**R**ating) пользователя u для товара i
- ▶ \hat{R}_{ui} – предсказанный рейтинг пользователя u для товара i

Коллаборативная фильтрация

- ▶ User-based
 - Определяем «похожесть» пользователей по их оценкам
- ▶ Item-based
 - Определяем «похожесть» товаров по их оценкам



Пример

		 Book 1	 Book 2	 Book 3	 Book 4	 Book 5
 User A						
 User B						
 User C						
 User D						

User-based filtering

Сходство двух пользователей u, v вычисляем через корреляцию Пирсона:

$$w_{uv} = \frac{\sum_{i \in I_{uv}} (R_{ui} - \bar{R}_u)(R_{vi} - \bar{R}_v)}{\sqrt{\sum_{i \in I_{uv}} (R_{ui} - \bar{R}_u)^2} \sqrt{\sum_{i \in I_{uv}} (R_{vi} - \bar{R}_v)^2}}$$

- I_{uv} - множество товаров, которые оценили пользователи u, v ;
- \bar{R}_u - среднее оценок пользователя u .

User-based filtering

- ▶ Находим множество пользователей $U(u)$, похожих на данного:

$$U(u) = \{v \in U \mid w_{uv} > \alpha\}$$

- ▶ Считаем как часто товар покупался пользователями из $U(u)$:

$$p_i = \frac{|\{u \in U(u) \mid \exists R_{ui}\}|}{|U(u)|}$$

- ▶ Прогноз оценки для товара i от пользователя u :

$$\hat{R}_{ui} = \bar{R}_u + \frac{\sum_{v \in U(u)} w_{uv} (R_{vi} - \bar{R}_v)}{\sum_{v \in U(u)} |w_{uv}|}$$

- ▶ Пользователю u рекомендуем k товаров с наибольшими p_i , либо \hat{R}_{ui} .

Пример

						
	5	?	1	1		2
		2		4		4
	4	5		1	1	2
			3	5	2	
	2		1		4	4

Item-based filtering

Сходство двух товаров i, j вычисляем через корреляцию Пирсона:

$$w_{ij} = \frac{\sum_{u \in U_{ij}} (R_{ui} - \bar{R}_i)(R_{uj} - \bar{R}_j)}{\sqrt{\sum_{u \in U_{ij}} (R_{ui} - \bar{R}_i)^2} \sqrt{\sum_{u \in U_{ij}} (R_{uj} - \bar{R}_j)^2}}$$

- U_{ij} - множество пользователей, которые оценили товары i, j ;
- \bar{R}_i - среднее оценок товара i .

Item-based filtering

- ▶ Находим множество товаров $I(i)$, похожих на данный:

$$I(i) = \{j \in I | w_{ij} > \alpha\}$$

- ▶ Находим максимально схожий товар, который уже покупался:

$$p_i = \max_j w_{ij}$$

- ▶ Прогноз оценки для товара i от пользователя u :

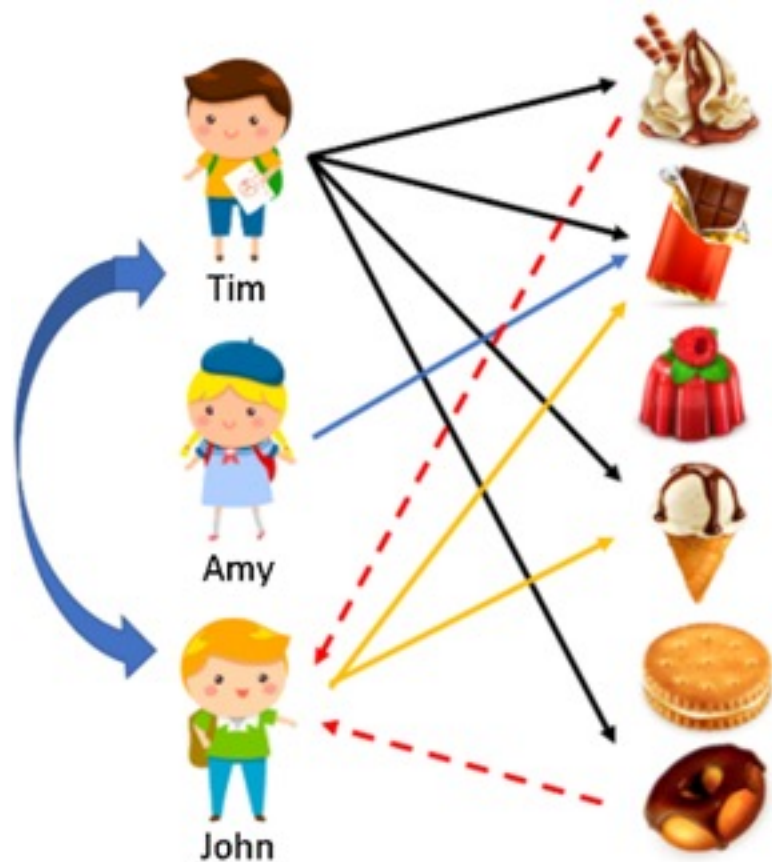
$$\hat{R}_{ui} = \frac{\sum_{j \in I(i)} w_{ij} R_{uj}}{\sum_{j \in I(i)} |w_{uj}|}$$

- ▶ Пользователю u рекомендуем k товаров с наибольшими p_i , либо \hat{R}_{ui} .

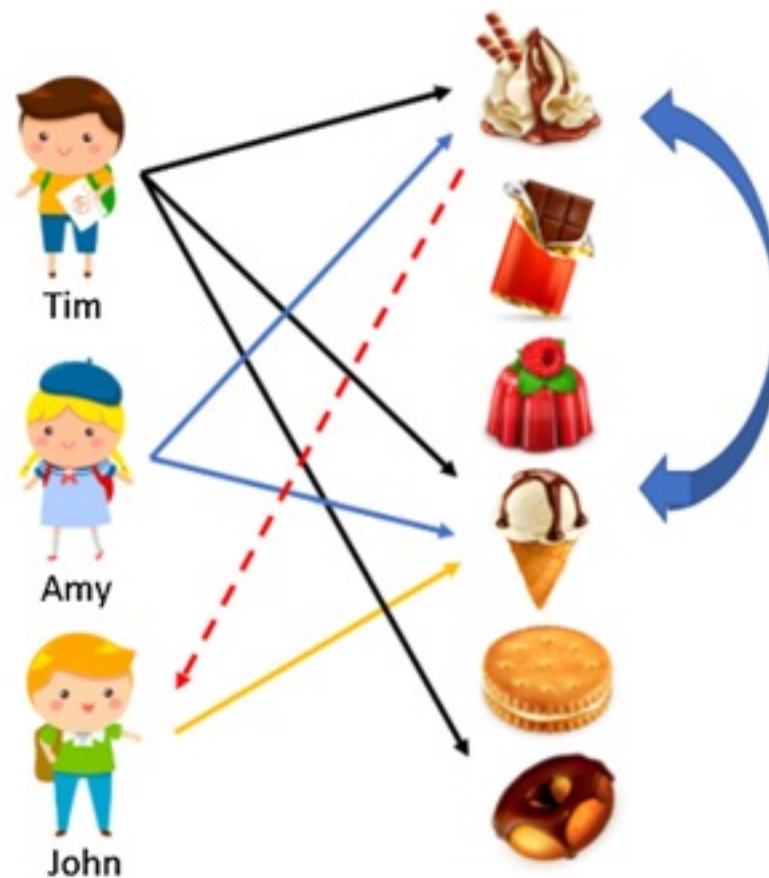
Пример

						
	5		1	1		2
		2		4		4
	4	5		1	1	2
			3	5	2	?
	2		1		4	4

Резюме



(a) User-based filtering



(b) Item-based filtering

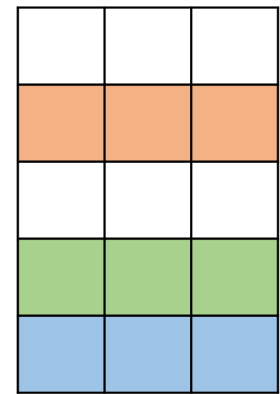
Модели со скрытыми переменными



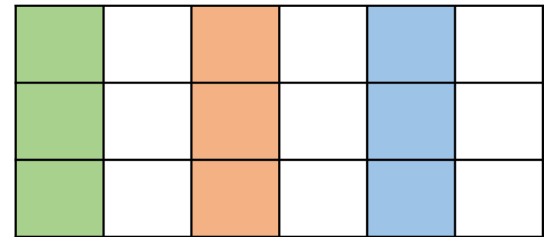
Идея



≈



×



Модель со скрытыми переменными

- ▶ Будем представлять матрицу оценок R через произведение двух матриц:

$$R = P^T Q$$

- P – матрица представлений пользователей;
- Q – матрица представлений товаров;

- ▶ Матрицы находим минимизируя ошибку прогноза:

$$\sum_{u,i} (R_{ui} - \bar{R}_u - \bar{R}_i - p_u^T q_i)^2 \rightarrow \min_{P,Q}$$

Latent Factor Model (LFM)

- ▶ Будем представлять матрицу оценок R через произведение двух матриц:

$$R = P^T Q$$

- P – матрица представлений пользователей;
- Q – матрица представлений товаров;

- ▶ Матрицы находим минимизируя ошибку прогноза:

$$\sum_{u,i} (R_{ui} - \bar{R}_u - \bar{R}_i - p_u^T q_i)^2 + \lambda \sum_{u \in U} \|p_u\|^2 + \mu \sum_{i \in I} \|q_i\|^2 \rightarrow \min_{P,Q}$$

Оптимизация SGD

- ▶ Значения матриц P и Q находим с помощью стохастического градиентного спуска:

$$p_{uk} = p_{uk} + \eta q_{ik}(R_{ui} - \bar{R}_u - \bar{R}_i - p_u^T q_i)$$

$$q_{ik} = q_{ik} + \eta p_{uk}(R_{ui} - \bar{R}_u - \bar{R}_i - p_u^T q_i)$$

Alternating Least Squares

- ▶ Альтернативный способ оптимизации.
- ▶ Зафиксируем матрицу Q . Тогда оптимальная P :

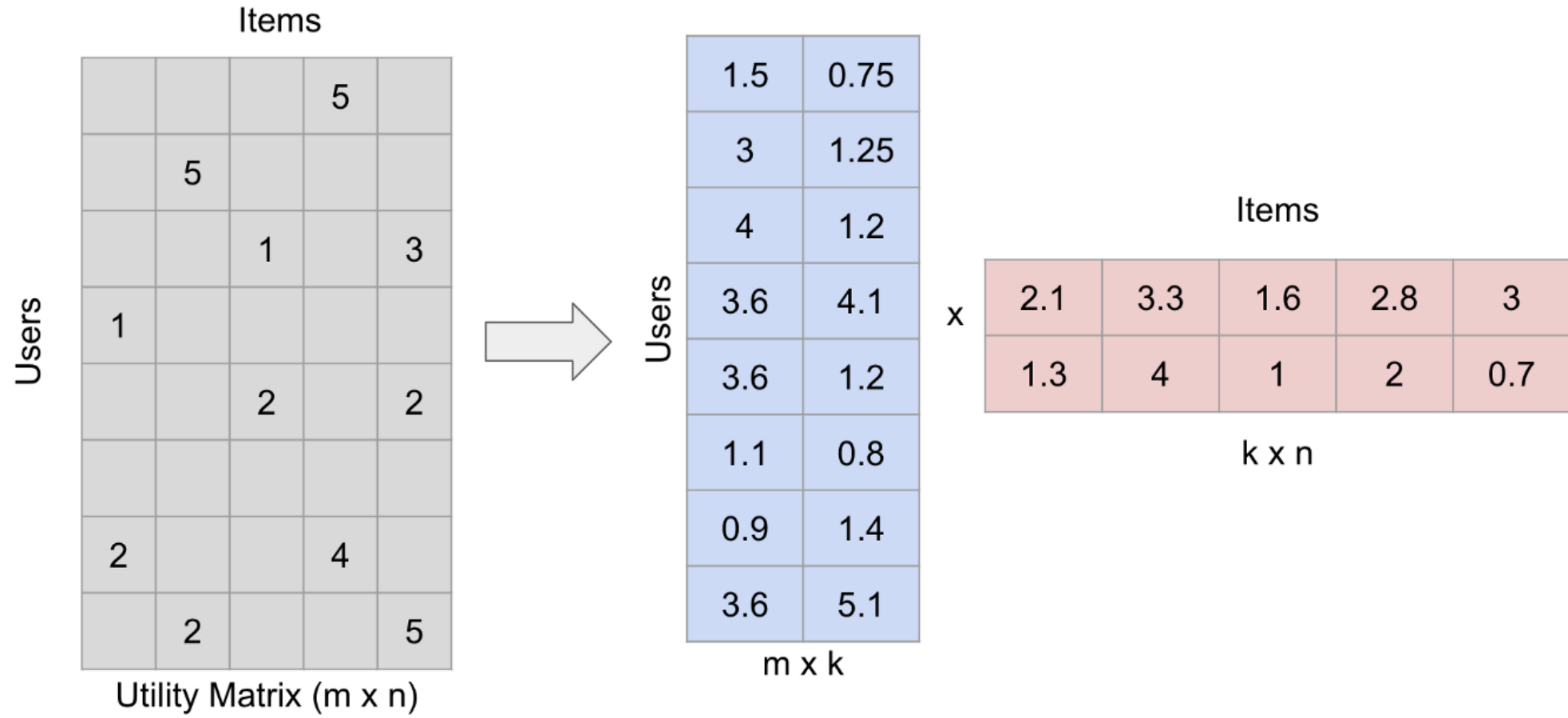
$$p_u = \left(\sum_i q_i q_i^T \right)^{-1} \sum_i r_{ui} q_i$$

- ▶ Затем, фиксируем матрицу P . Тогда оптимальная Q :

$$q_i = \left(\sum_u p_u p_u^T \right)^{-1} \sum_u r_{ui} p_u$$

- ▶ Повторяем шаги оптимизации.

Пример



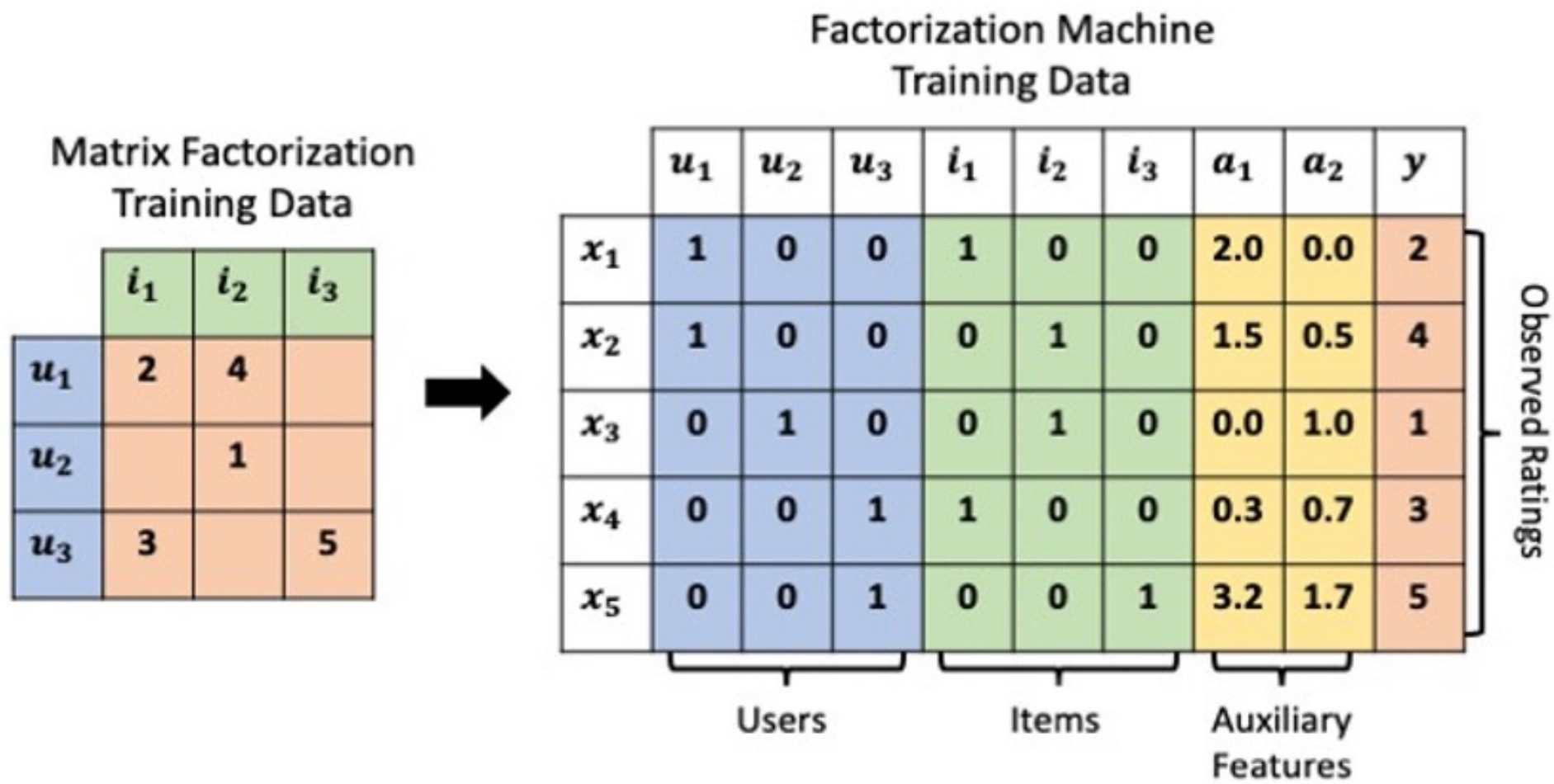
Factorization Machines



Мотивация

- ▶ Можно ли использовать дополнительную информацию о пользователях и товарах?
- ▶ Как это сделать?

Идея



Обозначения

Feature vector \mathbf{x}																	Target y					
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$\mathbf{x}^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$\mathbf{x}^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...		TI	NH	SW	ST	...		
	User				Movie					Other Movies rated					Time	Last Movie rated						

Модель

- ▶ Прогноз оценки пользователя $\hat{y}(x)$ для товара x :























$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n (v_i^T v_j) x_i x_j$$

- w_i, v_i – веса модели;
 - n – число пар оценок в данных.
- ▶ Обучаем модель минимизируя ошибку прогноза оценок.

Заключение



Резюме

						
		Book 1	Book 2	Book 3	Book 4	Book 5
	User A					
	User B					
	User C					
	User D					

Вопросы

- ▶ Опишите основные идеи, как работают следующие подходы к построению рекомендательных систем: user-based, item-based, подход на основе скрытых представлений.