Матричные разложения

обзор методов и применение в машинном обучении для понижения размерности, заполнения пропусков и рекомендаций

Матричные разложения

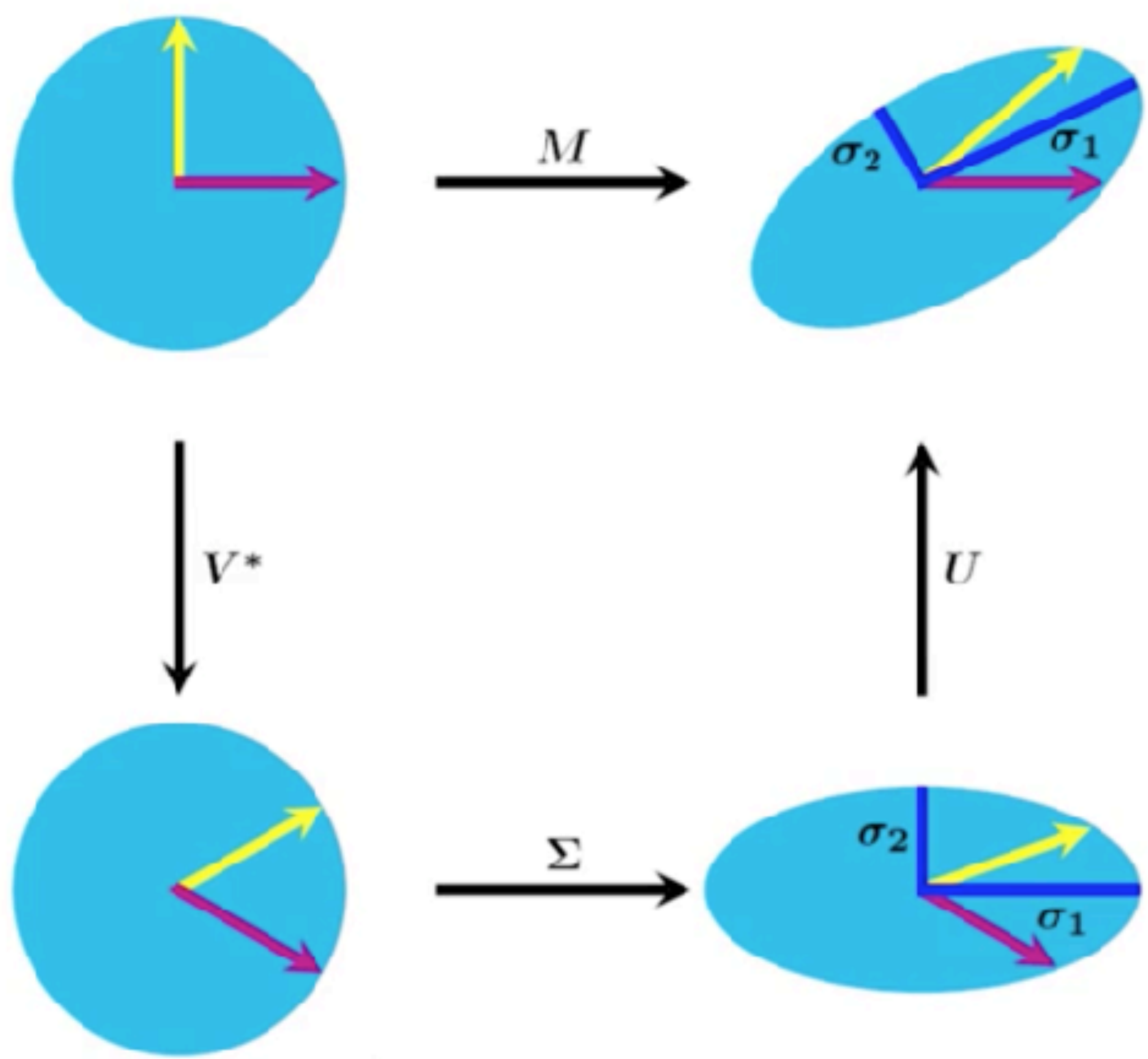
$$X \approx U \cdot V^T$$

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

$$\sum_{i,j}(x_{ij} - < u_i, v_j >)^2 \to \min$$

SVD в линейной алгебре

- $X = U \cdot \Sigma \cdot V^T$
- U, V — ортогональные, Σ — диагональная

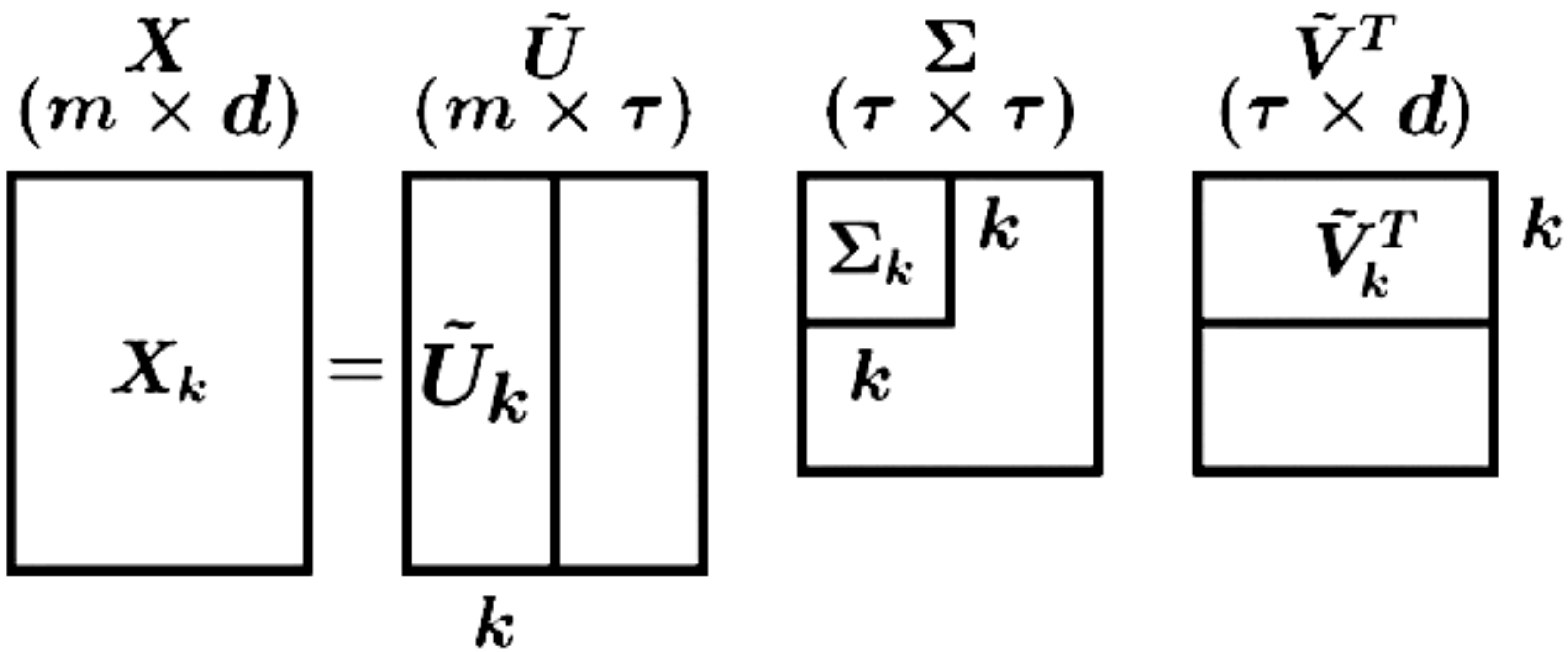


Приближение матрицы с помощью SVD

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

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

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



Приближение матрицы с помощью SVD

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

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

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

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

$$U = \tilde{U}_k \Sigma_k \qquad V = \tilde{V}_k$$

«SVD» в машинном обучении

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

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

- u_i — "профили" объектов
- v_j — "профили" исходных признаков

Матрица рейтингов и SVD

	Пила	Улица Вязов	Ванильное небо (j)	1+1	
Маша	5	4	1	2	
Юля	5	5	2		
Вова			3	5	i
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"

Градиентный спуск

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

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

- $\varepsilon_{ij} = (\langle u_i, v_j \rangle - x_{ij})$ — ошибка на x_{ij}

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

SGD

GD:

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

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

$\Rightarrow$

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

Метод чередующихся наименьших квадратов (ALS)

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

- Повторяем до сходимости:

$$\frac{\partial Q}{\partial u_i} = 0 \Rightarrow u_i \qquad \frac{\partial Q}{\partial v_j} = 0 \Rightarrow v_j$$

ALS

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

- $$\frac{\partial Q}{\partial u_i} = \sum_j 2(\langle u_i, v_j \rangle - x_{i,j}) v_j = 0$$
  $$\sum_j v_j \langle v_j, u_i \rangle = \sum_j x_{ij} v_j$$
- $$\sum_j v_j v_j^T u_i = \sum_j x_{ij} v_j$$
- $$\left(\sum_j v_j v_j^T \right) u_i = \sum_j x_{ij} v_j$$

ALS

- Повторяем по случайным i, j до сходимости:

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

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

(решение системы линейных уравнений)

Регуляризация

- $$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 \to \min_{u_i, v_j}$$
- α и β — небольшие положительные числа (0.001, 0.01, 0.5)

Постановка задачи рекомендации

j

		Пила	Улица Вязов	Ванильное небо	1+1
	Маша	5	4	1	2
	Юля	5	5	2	
	Вова			3	5
i	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»

Оптимизируемый функционал

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

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

Оптимизируемый функционал

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

$$+ \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 +$$

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

- Добавили сдвиг μ, базовые предикторы пользователя и фильма b

- L2-регуляризация

Рекомендации товаров

j

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

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

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

$$\Rightarrow$$

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

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

Explicit и implicit

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

Implicit матричное разложение

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

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

- $\mathbf{w}_{ij}$ принимает большие значения для $x_{ij} \neq 0$ и значительно меньшие для $x_{ij} = 0$

Implicit ALS

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

$$\mathbf{w}_{ij} = 1 + \alpha |x_{ij}| \qquad \alpha = 10, 100, 1000$$

u_i, v_j оцениваем с помощью ALS

Вероятностный подход

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

Распределение Пуассона и матричные разложения

- $x_{ij} \sim Poiss\langle u_i, v_j \rangle$
- $P\left(x_{ij}\right) = \frac{\langle u_i, v_j \rangle^{x_{ij}}}{x_{ij}!} e^{-\langle u_i, v_j \rangle}$
- $\prod\limits_{i,j} \frac{\langle u_i, v_j \rangle^{x_{ij}}}{x_{ij}!} e^{-\langle u_i, v_j \rangle} \to \max$
- $\sum\limits_{i,j} \langle u_i, v_j \rangle - x_{ij} \ln \langle u_i, v_j \rangle \ln x_{ij}! \to \min$
- $\sum\limits_{i,j} \langle u_i, v_j \rangle - x_{ij} \ln \langle u_i, v_j \rangle \to \min$

SGD для неотрицательного матричного разложения (NMF)

- $$Q = \sum_{i,j} \langle u_i, v_j \rangle - x_{ij} \ln \langle u_i, v_j \rangle \to \min$$
- $$\frac{\partial Q}{\partial u_i} = \sum_j v_j - \frac{x_{ij}}{\langle u_i, v_j \rangle} v_j =$$
$$= \sum_j \underbrace{\frac{\langle u_i, v_j \rangle - x_{ij}}{\langle u_i, v_j \rangle}}_{\tilde{\varepsilon}_{ij}\text{ - “относительная ошибка прогноза”}} v_j \to \min$$

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

Проблемы NMF

- Некорректно поставлена: если U_0, V_0 — решение, то $U = U_0 \cdot Y,\ V = V_0^T \cdot Y^{-T}$ тоже может быть решением

- $Q(X, UV^T)$ не выпукла по совокупности аргументов (U, V), не можем использовать методы, которые находят глобальный минимум

Блочно-покоординатные методы минимизации

- Вход: $U_0 \geq 0, V_0 \geq 0$
- Цикл
  - $V \leftarrow f(X, U, V)$
  - $U \leftarrow g(X, U, V)$
- $f(X, U, V) = g^T(X^T, V^T, U^T)$, поскольку задача симметрична

NMF с нормой Фробениуса

- Оптимизационная задача:

$$\min_{U \geq 0, V \geq 0} ||X - UV^T||_F^2 \Rightarrow (U*, V*)$$

- Без ограничения неотрицательности — SVD
- Базовый метод — поочерёдный градиентный спуск.

Мультипликативные обновления

- Идея: выбрать шаги GD так, чтобы обновления стали мультипликативными

$$\frac{\partial Q}{\partial v_{jk}} = \sum_{i=1}^{n} u_{ik}\hat{x}_{ij} - \sum_{i=1}^{n} u_{ik}x_{ij}, \ \ \eta_{jk} = \frac{v_{jk}}{\sum u_{ik}\hat{x}_{ij}}$$

$$v_{jk}^{(t+1)} = v_{jk}^{(t)} - \eta_{jk}\left(\sum_{i=1}^{n} u_{ik}\hat{x}_{ij} - \sum_{i=1}^{n} u_{ik}x_{ij}\right)$$

$$= v_{jk} \cdot \frac{\sum u_{ik}x_{ij}}{\sum u_{ik}\hat{x}_{ij}}$$

- В матричном виде: $V \leftarrow V \otimes (X^T U) \oslash (\hat{X}^T U)$

ALS и ANLS

1. На каждом шаге находится решение задачи LS по одной из компонент и проецируется на неотрицательную область.

   - Быстрый, но грубый: итерационный процесс не сходится, функция потерь осциллирует.

2. На каждом шаге точно находится покомпонентный минимум в неотрицательной области

   - Точный, но медленный.

HALS

- На каждом шаге точно находится минимум по столбцу u_k или строке v_k
    - на каждом шаге небольшие вычислительные затраты
    - сходится быстрее мультипликативных обновлений
    - Чувствителен к начальному приближению