

Сингулярное разложение

Виктор Китов

victorkitov.github.io

Курс поддержан
фондом
'Интеллект'



$$\begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix} = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix} \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix} \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix}$$

Победитель
конкурса VK среди
курсов по IT



Содержание

- 1 Определение сингулярного разложения
- 2 Сокращенное сингулярное разложение
- 3 Применения сингулярного разложения
- 4 Простейшая рекомендательная система

Сингулярное разложение

Сингулярное разложение (singular value decomposition, SVD):

Каждая матрица $X \in \mathbb{R}^{N \times D}$, $\text{rank } X = R$, может быть разложена:

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

где

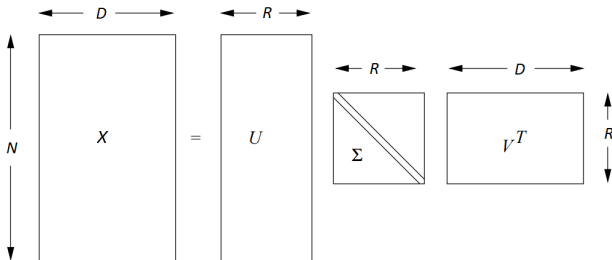
- $U \in \mathbb{R}^{N \times R}$, $\Sigma \in \mathbb{R}^{R \times R}$, $V^T \in \mathbb{R}^{R \times D}$
- $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_R\}$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R \geq 0$
- $U^T U = I$, $V^T V = I$, $I \in \mathbb{R}^{R \times R}$ - единичная матрица.

Эквивалентно:

$$X = \sum_{i=1}^R u_i \sigma_i v_i^T$$

где u_i - i -ая колонка U and v_i^T - i -ая строка V^T .

Интерпретация сингулярного разложения



- Столбцы U - ортонормированный базис столбцов X
- Строки V^T - ортонормированный базис строк X
- Σ - важности базисных векторов.

SVD дает компактное представление низкоранговой матрицы.

Столбцы V = главные компоненты X

$$X^T X = (U \Sigma V^T)^T U \Sigma V^T = (V \Sigma U^T) U \Sigma V^T = V \Sigma^2 V^T$$

Домножая на V , получим¹

$$X^T X V = V \Sigma^2 V^T V = V \Sigma^2 \quad (1)$$

- V состоит из СВ $X^T X$, отвечающих СВ $\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2$ - это R главных компонент.
- U - коэффициенты разложения объектов-строк X по главным компонентам.

¹Сингулярные значения X - корень из СВ $X^T X$, равные $\sigma_1, \dots, \sigma_R$.

Нахождение U

$$XX^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T$$

Домножая справа на U , получим

$$XX^T U = U\Sigma^2 U^T U = U\Sigma^2.$$

U состоит из СВ XX^T , отвечающих СВ $\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2$.

SVD: существование & единственность

Теорема 1

Для $\forall X \in \mathbb{R}^{N \times D}$ сингулярное разложение существует.

Теорема 2

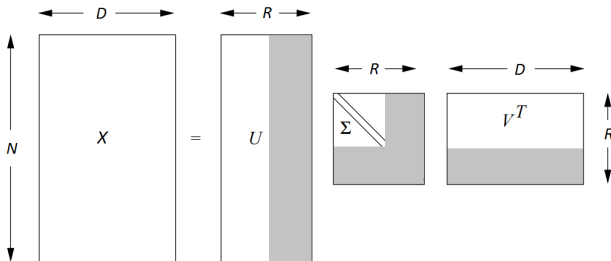
Сингулярное разложение единственно с точностью до знака $\Leftrightarrow X^T X \in \mathbb{R}^{D \times D}$ содержит D уникальных СВ.

С точностью до знака означает, что мы всегда можем одновременно поменять знаки u_i и v_i^T для $\forall i = 1, 2, \dots, R$.

Содержание

- 1 Определение сингулярного разложения
- 2 Сокращенное сингулярное разложение
- 3 Применения сингулярного разложения
- 4 Простейшая рекомендательная система

Сокращенное сингулярное разложение



Сокращенное сингулярное разложение порядка K (truncated SVD):

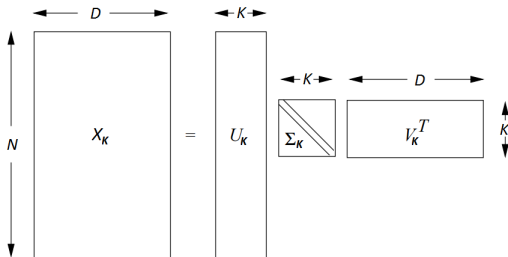
убрать наименее важные столбцы U и строки V^T .

- Важность измеряется $\sigma_1, \sigma_2, \dots, \sigma_R$.

$$\text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\} \longrightarrow \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, 0, 0, \dots, 0\}$$

$$X = \sum_{i=1}^R u_i \sigma_i v_i^T \approx \sum_{i=1}^K u_i \sigma_i v_i^T$$

Сокращенное сингулярное разложение



Упрощение до ранга $K \leq R$:

$$\hat{X} = U_K \Sigma_K V_K \approx X$$

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\} \longrightarrow \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K\} = \Sigma_K$$

$$U = [u_1, u_2, \dots, u_K, u_{K+1}, \dots, u_R] \longrightarrow [u_1, u_2, \dots, u_K] = U_K$$

$$V = [v_1, v_2, \dots, v_K, v_{K+1}, \dots, v_R] \longrightarrow [v_1, v_2, \dots, v_K] = V_K$$

Свойства сокращенного сингулярного разложения

Норма Фробениуса для матриц

$$\|X\|_F^2 = \sum_{n=1}^N \sum_{d=1}^D x_{nd}^2$$

- Для матрицы X и её аппроксимации \hat{X} :

$$\text{ошибка аппроксимации} = \|\hat{X} - X\|_F^2$$

Теорема 3

Пусть $X \in \mathbb{R}^{N \times D}$ аппроксимируется $\hat{X} = U_K \Sigma_K V_K$. Тогда:

- 1 $\text{rank } \hat{X} = K$.
- 2 $\hat{X} = \arg \min_{B: \text{rank } B \leq K} \|X - B\|_F^2$

Выбор порядка аппроксимации K

Теорема 4

Для \forall матрицы X и её разложения $X = U\Sigma V^T$,
 $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$:

$$\|X\|_F^2 = \sum_{i=1}^R \sigma_i^2$$

$$X = U\Sigma V^T \quad \Sigma = \text{diag}\{\sigma_1, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\}$$

$$\hat{X} = U\Sigma_K V^T \quad \Sigma_K = \text{diag}\{\sigma_1, \dots, \sigma_K, 0, 0, \dots, 0\}$$

$$X - \hat{X} = U(\Sigma - \Sigma_K) V^T \quad \Sigma - \Sigma_K = \text{diag}\{0, 0, \dots, \sigma_{K+1}, \dots, \sigma_R\}$$

$$\|X - \hat{X}\|_F^2 = \sum_{i=K+1}^R \sigma_i^2$$

Выбор порядка аппроксимации K

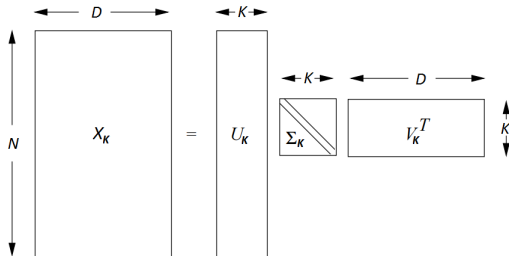
Используя теорему 4, выберем K , дающую относительную ошибку меньше порога:

$$K = \arg \min_K \left\{ \frac{\|X - \hat{X}\|_F^2}{\|X\|_F^2} = \frac{\sum_{i=K+1}^R \sigma_i^2}{\sum_{i=1}^R \sigma_i^2} < \text{threshold} \right\}$$

Содержание

- 1 Определение сингулярного разложения
- 2 Сокращенное сингулярное разложение
- 3 Применения сингулярного разложения**
- 4 Простейшая рекомендательная система

Снижение размерности



- строки U дают компактное представление объектов X .
- $x_n \in \mathbb{R}^D \longrightarrow u_n \in \mathbb{R}^K, \quad K \leq D$.

Пример: сжатие чёрно-белых изображений²

(a) 5 singular Vectors



(b) 10 singular Vectors



(c) 25 singular Vectors



(c) 35 singular Vectors



(c) 50 singular Vectors



(d) Full image



Пример: сжатие цветных изображений³

Rank= 5



Rank= 10



Rank= 25



Rank= 35



Rank= 50



Rank=187



³Сжатие - независимо по R,G,B каналам.

Экономия памяти

Рассчитайте стоимость хранения $X \in \mathbb{R}^{N \times D}$, предполагая $N \geq D$:

представление X	требования по памяти
исходная X	?
полностью сингулярно разложенная X	?
сокращенно синг. разложенная ранга K	?

Вычислительные затраты

- Умножение Xq
 - X - нормализованное представление документов
 - q - нормализованный поисковый запрос

представление X	сложность Xq
исходная X	?
полностью сингулярно разложенная X	?
сокращенно синг. разложенная ранга K	?

Нахождение похожих объектов и похожих признаков

- Похожие объекты имеют похожие признаки.
- Похожие признаки встречаются в объектах.
- Пример: обработка текстов.
 - сингулярное разложение дает высокоуровневое семантическое представление документов
 - можем сравнивать документы на семантическом уровне
 - синонимы объединяются
 - можем сравнивать слова

Содержание

- 1 Определение сингулярного разложения
- 2 Сокращенное сингулярное разложение
- 3 Применения сингулярного разложения
- 4 Простейшая рекомендательная система

Построение рекомендаций фильмов

	Терминатор	Гладиатор	Рэмбо	Титаник	История любви	Спеши любить
Андрей	4	5	5	0	0	0
Иван	4	4	5	0	0	0
Сергей	5	5	4	0	0	0
Анна	0	0	0	5	5	5
Мария	0	0	0	5	5	4
Наталья	0	0	0	4	5	4

Сингулярное разложение

$$U = \begin{pmatrix} 0. & 0.6 & -0.3 & 0. & 0. & -0.8 \\ 0. & 0.5 & -0.5 & 0. & 0. & 0.6 \\ 0. & 0.6 & 0.8 & 0. & 0. & 0.2 \\ 0.6 & 0. & 0. & -0.8 & -0.2 & 0. \\ 0.6 & 0. & 0. & 0.2 & 0.8 & 0. \\ 0.5 & 0. & 0. & 0.6 & -0.6 & 0. \end{pmatrix}$$

$$\Sigma = \text{diag}\{(14. \quad 13.7 \quad 1.2 \quad 0.6 \quad 0.6 \quad 0.5)\}$$

$$V^T = \begin{pmatrix} 0. & 0. & 0. & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.6 & 0.6 & 0. & 0. & 0. \\ 0.5 & 0.3 & -0.8 & 0. & 0. & 0. \\ 0. & 0. & 0. & -0.2 & 0.8 & -0.6 \\ -0. & -0. & -0. & 0.8 & -0.2 & -0.6 \\ 0.6 & -0.8 & 0.2 & 0. & 0. & 0. \end{pmatrix}$$

Сокращенное сингулярное разложение (K=2)

$$U_2 = \begin{pmatrix} 0. & 0.6 \\ 0. & 0.5 \\ 0. & 0.6 \\ 0.6 & 0. \\ 0.6 & 0. \\ 0.5 & 0. \end{pmatrix}$$

$$\Sigma_2 = \text{diag}\{(14. \quad 13.7)\}$$

$$V_2^T = \begin{pmatrix} 0. & 0. & 0. & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.6 & 0.6 & 0. & 0. & 0. \end{pmatrix}$$

Перешли в на семантический уровень "тем"

- темы среди фильмов - боевик / мелодрама
- темы среди людей - мужчины / женщины

Построение рекомендаций

- Требуется построить рекомендации фильмов для нового человека:

$$x = (5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$$

- 1 Отображаем x в пространство тем фильмов y (снижение размерности):

$$y = V_2^T x = (0 \quad 2.7)$$

- 2 Построение рекомендаций: отображаем y в исходное пространство всех оценок:

$$\hat{x} = yV_2^T = (1.5 \quad 1.6 \quad 1.6 \quad 0 \quad 0 \quad 0)$$

Заключение

- Сингулярное разложение $X = U\Sigma V^T$, $U^T U = I$, $V^T V = I$, $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$ существует $\forall X$.
- Сокращенное сингулярное разложение порядка K
 - решает задачу: $\hat{X} = \arg \min_{B: \text{rank } B \leq K} \|X - B\|_F^2$
 - извлекает тематическую структуру объектов и признаков
 - разложение по темам=главным компонентам: снижение размерности
 - сокращает
 - расходы по памяти / на вычисления
 - позволяет построить простую рекомендательную систему
 - недостаток: отсутствие оценки=0 трактуется как оценка.