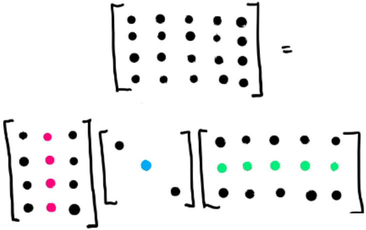
Сингулярное разложение

Виктор Китов

victorkitov.github.io



Курс поддержан фондом 'Интеллект'

Победитель конкурса VK среди курсов по IT

Содержание

Сингулярное разложение

Сингулярное разложение (singular value decomposition, SVD):

Каждая матрица $X \in \mathbb{R}^{NxD}$, $\operatorname{rank} X = R$, может быть разложена:

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

где

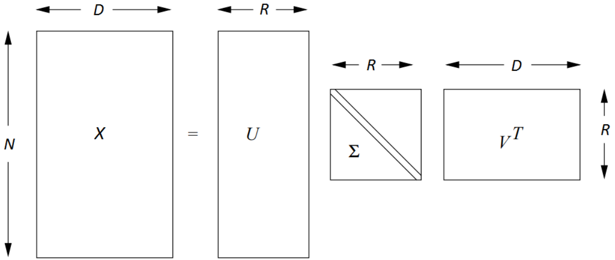
- $U \in \mathbb{R}^{NxR}$, $\Sigma \in \mathbb{R}^{RxR}$, $V^T \in \mathbb{R}^{RxD}$
- $\Sigma = diag\{\sigma_1, \sigma_2, ...\sigma_R\}$, $\sigma_1 \ge \sigma_2 \ge ... \ge \sigma_R \ge 0$ - сингулярные числа X.
- $U^T U = I$, $V^T V = I$, $I \in \mathbb{R}^{RxR}$ - единичная матрица.

Эквивалентно:

$$X = \sum_{i=1}^{R} \mathbf{u}_i \sigma_i \mathbf{v}_i^T$$

где u_i - i-й столбец U, а 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

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

Столбцы V=главные компоненты X

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

Домножая на V, получим

$$X^T X V = V\Sigma^2 V^T V = V\Sigma^2 \tag{1}$$

Столбцы V=главные компоненты X

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

Домножая на V, получим

$$X^T X V = V\Sigma^2 V^T V = V\Sigma^2 \tag{1}$$

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

Нахождение U

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

Домножая справа на U, получим

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

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

SVD: существование & единственность

Теорема 1

Для $\forall X \in \mathbb{R}^{NxD}$ *сингулярное разложение существует.*

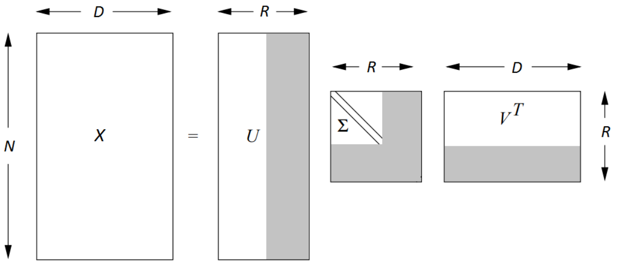
Теорема 2

Сингулярное разложение единственно с точностью до знака <=> $X^T X \in \mathbb{R}^{DxD}$ *содержит* D *уникальных СВ.*

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

Содержание

Сокращенное сингулярное разложение



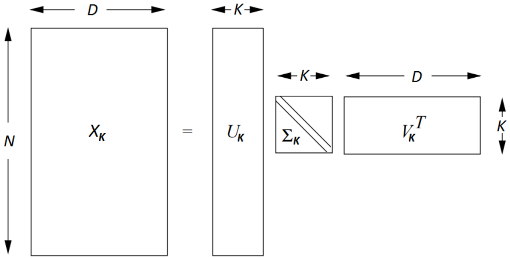
Сокращенное сингулярное разложение порядка K (truncated SVD): убрать наименее важные столбцы U и строки V^T.

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

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

$$X = \sum_{i=1}^{R} \mathbf{u}_i \sigma_i \mathbf{v}_i^T \approx \sum_{i=1}^{K} \mathbf{u}_i \sigma_i \mathbf{v}_i^T$$

Сокращенное сингулярное разложение



Упрощение до ранга $K \le R$:

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

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, ...\sigma_K, \sigma_{K+1}, ...\sigma_R\} \to \text{diag}\{\sigma_1, \sigma_2, ...\sigma_K\} = \Sigma_K$$
$$U = [u_1, u_2, ...u_K, u_{K+1}, ...u_R] \longrightarrow [u_1, u_2, ...u_K] = U_K$$
$$V = [v_1, v_2, ...v_K, v_{K+1}, ...v_R] \longrightarrow [v_1, v_2, ...v_K] = V_K$$

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 $\widehat{X}$:

$$\text{ошибка аппроксимации} = \left\|\widehat{X} - X\right\|_F^2$$

Теорема 3

Пусть $X \in \mathbb{R}^{NxD}$ *аппроксимируется* $\widehat{X} = U_K \Sigma_K V_K$. *Тогда:*

1. $\operatorname{rank} \widehat{X} = K$.
2. $\widehat{X} = \arg\min_{B:\operatorname{rank} B \le K} \|X - B\|_F^2$

Выбор порядка аппроксимации K

Теорема 4

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

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

$$\begin{aligned}
X &= U\Sigma V^T & \Sigma &= \mathrm{diag}\{\sigma_1, ...\sigma_K, \sigma_{K+1}, ...\sigma_R\} \\
\widehat{X} &= U\Sigma_K V^T & \Sigma_K &= \mathrm{diag}\{\sigma_1, ...\sigma_K, 0, 0, ...0\} \\
X - \widehat{X} &= U\left(\Sigma - \Sigma_K\right)V^T & \Sigma - \Sigma_K &= \mathrm{diag}\{0, 0, ...\sigma_{K+1}, ...\sigma_R\}
\end{aligned}$$

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

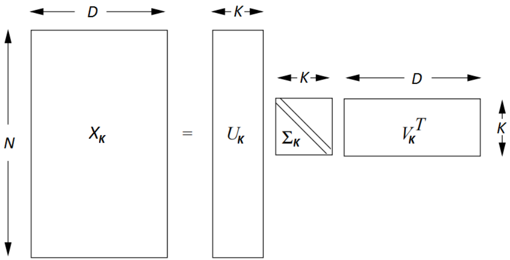
Выбор порядка аппроксимации K

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

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

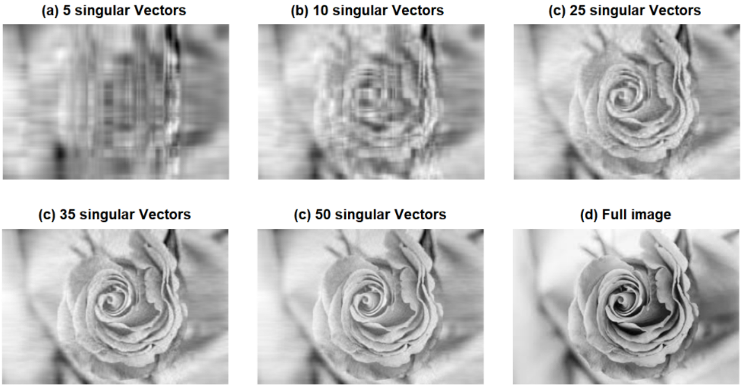
Содержание

Снижение размерности



- строки $(U_K \Sigma_K)$ - компактное представление объектов X.
    - K-мерное разложение по главным компонентам
- Также существуют
    - неотрицательные матричные разложения (non-negative matrix factorization)
    - стохастические разложения (PLSA, LDA)

Пример: сжатие чёрно-белых изображений[1]



[1] Первоисточник.

Пример: сжатие цветных изображений[2]

[2] Сжатие - независимо по R,G,B каналам.

Экономия памяти

Рассчитайте стоимость хранения $X \in \mathbb{R}^{NxD}$, 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.
- Пример: обработка текстов.
    - сингулярное разложение дает высокоуровневое семантическое представление документов
    - можем сравнивать документы на семантическом уровне
        - синонимы объединяются
    - можем сравнивать слова (столбцы V^T) по совстречаемости в документах!

Содержание

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

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

$$\Sigma = \mathsf{diag}\{\begin{pmatrix} 14. & 13.7 & 1.2 & 0.6 & 0.6 & 0.5 \end{pmatrix}\}$$

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

Сокращенное сингулярное разложение (K=2)

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

$$\Sigma_2 = \text{diag}\{\begin{pmatrix} 14. & 13.7 \end{pmatrix}\}$$

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

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:

$$x = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

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

$$y = V_2^T x = \begin{pmatrix} 0 & 2.7 \end{pmatrix}$$

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

$$\widehat{x} = yV_2^T = \begin{pmatrix} 1.5 & 1.6 & 1.6 & 0 & 0 & 0 \end{pmatrix}$$

Заключение

- Сингулярное разложение $X = U\Sigma V^T$, $U^T U = I$, $V^T V = I$, $\Sigma = \mathrm{diag}\{\sigma_1, ...\sigma_R\}$ существует $\forall X$.
- Сокращенное сингулярное разложение порядка K
    - решает задачу:$\widehat{X} = \arg\min_{B:\mathrm{rank}\, B \le K} \|X - B\|_F^2$
    - 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.