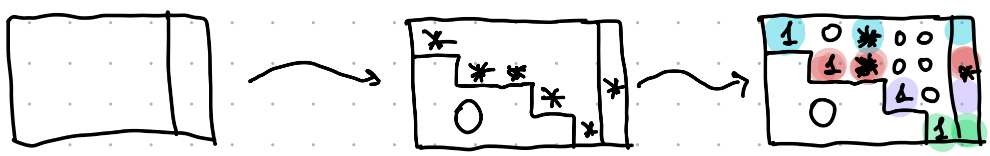
Матричные разложения

- → LU разложение (р-е Холецкого)
- → QR разложение
- → SVD разложение

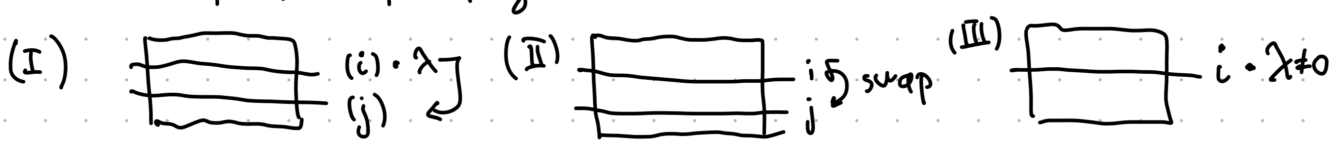
$X = AB$ матричная факторизация

① LU разложение

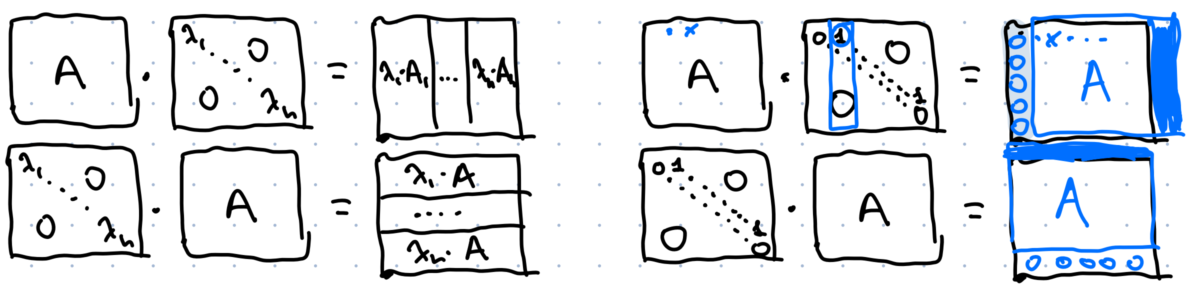
$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ \dots\dots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m \end{cases} \qquad \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right)$$



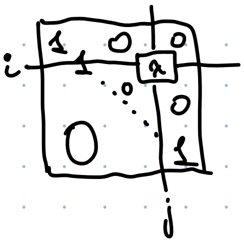
Элементарные преобразования



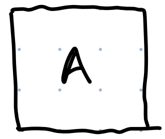
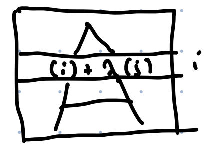
Упражнение



Умножение слева – воздействие на строки, справа – на столбцы

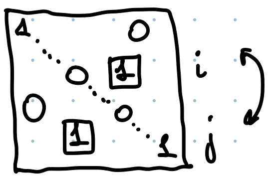
(I) $S_{ij}(\lambda) =$ 

$i \neq j$

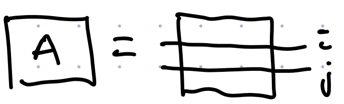
$S_{ij}(\lambda) \cdot$  $=$ 

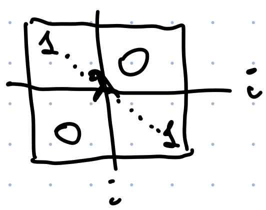
Все строки кроме i-ой прежние

$(i) \mapsto (i) + \lambda (j)$

(II) $P_{ij} =$  swap

В I_n поменяли строки i и j местами

$P_{ij} \cdot$  swap

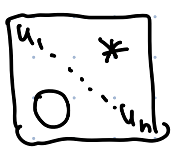
(III) $D_i(\lambda) =$ 

$D_i(\lambda) A$ – строка i $\times \lambda$

Алгоритм Гаусса

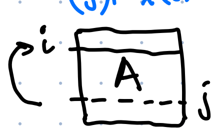
Начиная с этого момента $A_{[n \times n]}$

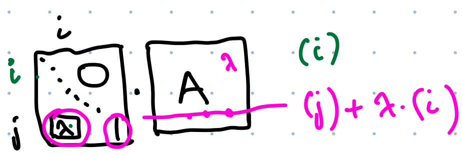
$A \to A_1 \to A_2 \to \ldots \to U$ – верхнетреугольная



Всегда к более нижней строке прибавлять более верхнюю

$(j) + \lambda(i)$

$\lambda \times$ 



$$U = \underbrace{S_n \cdot S_{n-1} \cdot \ldots \cdot S_1}_{L} \cdot \underset{P}{\uparrow} A$$

можно один разок переставить строки

$L^{-1} U = A$

Утв.

При умножении 2-х верхне Δ матриц всегда будет верхне Δ матрица

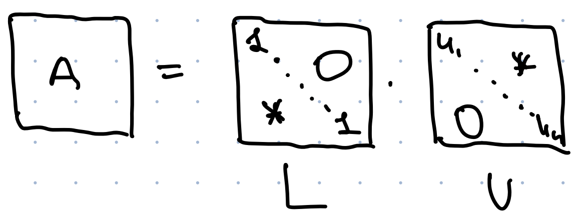
$$\hat{L}\hat{L}^{-1} = I$$

$$A = LU$$

$$U = \cancel{S_n}\,\cancel{S_{n-1}}\cdot\ldots S_1\cdot A$$

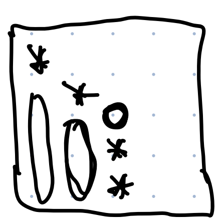
$$L^{-1} = L$$

$$\underbrace{S_1^{-1}\ldots S_{n-1}^{-1}S_n^{-1}}U = A$$

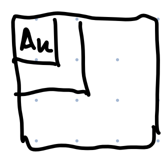


LU-разложение $\exists$ не всегда

Мерзость:



Условие существования:



$\det A_k \neq 0$

$\det A \neq 0$

Нафига нужно?

→ решать системы

$Ax = b$

$L\underbrace{Ux}_{y} = b$

$Ly = b$ прямой ход Гаусса

$Ux = y$ обратный ход Гаусса

Упражнение

(0) Сортировка строк чтобы наверху было самое большое числа

$$A = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 0 & 3 \\ 0 & 1 & 1 \end{pmatrix}$$

(1) Делим 1 столбец на 1ый диаг эл.
1-ую строку беру в U

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U = \begin{pmatrix} 3 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(2) $$A - LU = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 0 & 3 \\ 0 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4/3 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -4/3 & 3 \\ 0 & 1 & 1 \end{pmatrix}$$

повторяю (1) и (2)

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 1 & 0 \\ 0 & -3/4 & 0 \end{pmatrix} \quad U = \begin{pmatrix} 3 & 2 & 0 \\ 0 & -4/3 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

$$A - LU = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 13/4 \end{pmatrix}$$

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 1 & 0 \\ 0 & -3/4 & 1 \end{pmatrix} \quad U = \begin{pmatrix} 3 & 2 & 0 \\ 0 & -4/3 & 3 \\ 0 & 0 & 13/4 \end{pmatrix}$$

Сложность поиск LU: $O(n^3)$

Дальше каждая система решается за $O(n^2)$

$A x = \underset{\sim}{b}$

$A = LU$

$PA = LU$

Разложение Холецкого

LU-разложение для симметричной полож. опр. матрицы A

$A^T = A$

$\forall h \in \mathbb{R}^n \quad h^T A h > 0$

$h \neq 0$

→ Оно всегда $\exists$

→ вычисляется устойчиво

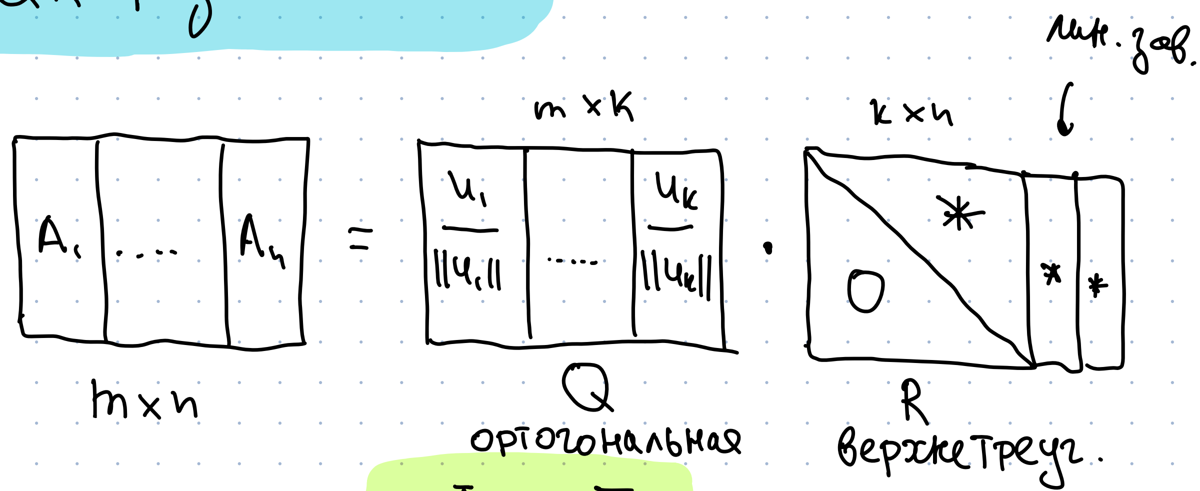
→ Все диаг эл. > 0 => единственное

$A = RR^T$ $\quad R$ – нижне-Δ

R – „квадратный корень" матрицы

$U^T U$ $\quad U$ – верхне Δ $\quad U^T = R$

② QR-разложение



$Q^T Q = I$

→ Оно определено $\forall$ прямоуг.
→ Используется внутри SVD
→ Если A кв. необр. тогда R тоже необр.

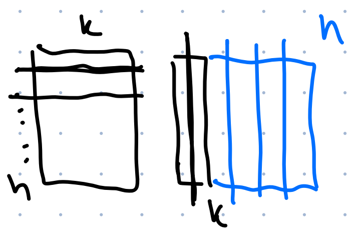
Упражнение

$y = Xw$ MSE $X\ [n \times k]$

$$\hat{w} = (X^T X)^{-1} X^T y$$

$\underset{[n\times k]\ [k\times n]}{X^T X}$ $O(k n^2)$ $\underset{[n\times n]}{(X^T X)^{-1}}$ $O(n^3)$ $O(n^{2.7\ldots})$

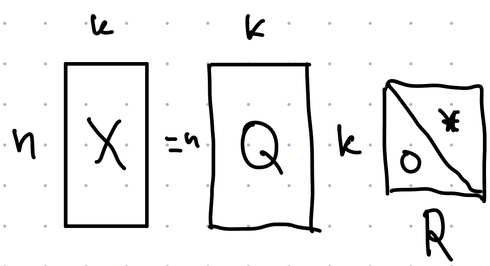
$\underset{[n\times k][k\times 1]}{X^T y}$ $O(kn)$



$$O(kn^2 + \cancel{kn} + n^{2.7\ldots})$$

$$O(kn^2 + n^{2.7\ldots})$$

$X = QR$ $\boxed{O(kn^2)}$



$rk(X) = k$

$\Downarrow$

$$(R^T R)^{-1} = R^{-1} R^{-T}$$

$$\hat{w} = (X^T X)^{-1} X^T y$$

$$\hat{w} = ((QR)^T (QR))^{-1} (QR)^T y$$

$$(R^T \underbrace{Q^T Q}_{I} R)^{-1} R^T Q^T y$$

$$R^{-1} R^{-T} R^T Q^T y$$

$$R\,\underbrace{R\,R}_{I}\,Q\,y$$

$$\hat{w} = R^{-1} Q^T y$$

$$R\hat{w} = Q^T y \quad \Rightarrow \quad O(n^2)$$

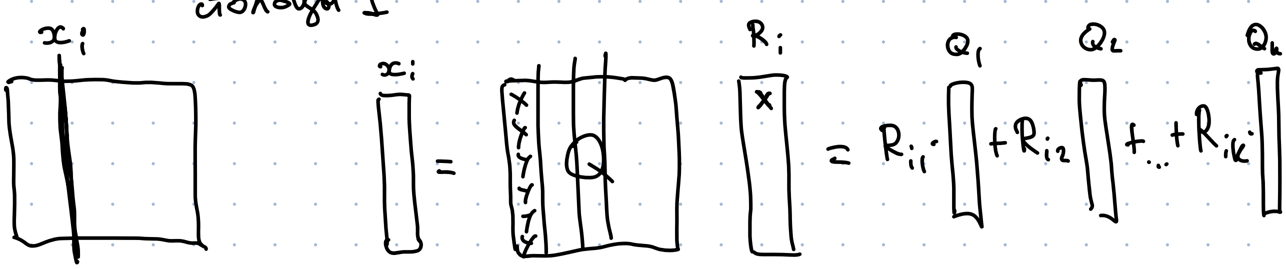
$\uparrow$ верхне △ и квадратная

Решение работает стабильнее, чем обращение матриц, но всё ещё хуже SGD

$$X = Q\,R \leftarrow$$

веса с которыми их можно сложить в столбцы из x

Q: столбцы $\perp$



Ортогонализация Грамма-Шмидта

$v_1, \dots, v_k \in V \qquad \mathbb{R}^n$

$\lambda_1 v_1 + \dots + \lambda_k v_k = 0$

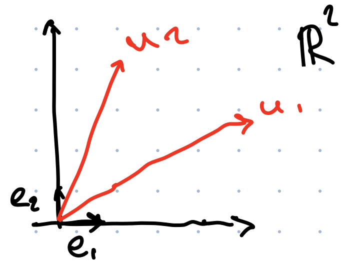
$\lambda_i = 0 \Rightarrow$ лин. нез.

$\exists\, \lambda_i \neq 0 \Rightarrow$ лин. зав.

$$\begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array}$$
$$\begin{pmatrix} 3 & 4 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 1 & 0 & 0 & 2 \end{pmatrix}$$

базис

ортогональный базис



$$1 + 2x + 4x^3 - 5x^4$$

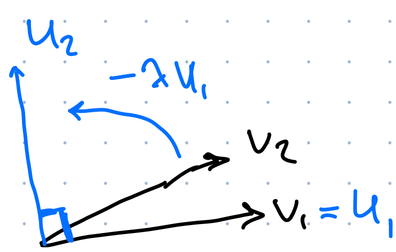
$$\begin{pmatrix} 1 \\ 2 \\ 4 \\ 5 \end{pmatrix} \qquad 1, x, x^3, x^4$$

$$\mathbb{R}^n + \langle \cdot ; \cdot \rangle \qquad x^T y$$

$$\cos(x, y) = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}$$

$$\langle x, y \rangle = 0 \Rightarrow x \perp y$$

ищем ортогональный базис:



$$u_1 = v_1$$

$$u_2 = v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} \cdot u_1$$

$$u_3 = v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} \cdot u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} \cdot u_2$$

$$\begin{cases} u_2 = v_2 - \lambda u_1 \\ u_2 \perp u_1 \end{cases}$$

. . . . .

$$u_k = v_k - \sum_{i=1}^{k} \frac{\langle v_k, u_i \rangle}{\langle u_i, u_i \rangle} \cdot u_i$$

$$0 = \langle u_2, u_1 \rangle = \langle v_2 - \lambda u_1, u_1 \rangle = $$
$$= \langle v_2, u_1 \rangle - \lambda \langle u_1, u_1 \rangle$$
$$\Rightarrow \quad \lambda = \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle}$$

$$\begin{cases} u_3 - v_3 - \lambda_1 u_1 - \lambda_2 u_2 \\ u_3 \perp u_2 \quad u_3 \perp u_1 \end{cases}$$

перпенд.

$$\begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \dots & u_k \end{bmatrix} \cdot \begin{bmatrix} 1 & \frac{\langle , \rangle}{\langle , \rangle} & & \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

$$\begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{u_1}{\|u_1\|} & \frac{u_2}{\|u_2\|} & \dots & \frac{u_k}{\|u_k\|} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \|u_1\| & & * \\ & \|u_2\| & \\ & & \ddots \\ 0 & & & \|u_k\| \end{bmatrix}}_{R}$$

$\|x\| = \sqrt{\langle x, x\rangle}$

$$\frac{\langle v_k, u_i\rangle}{\sqrt{\langle u_i, u_i\rangle}} = \frac{\langle v_k, u_i\rangle}{\|u_i\|}$$

Есть более эфф. алгоритмы поиска QR
- → метод вращений
- → метод отражений

③ SVD

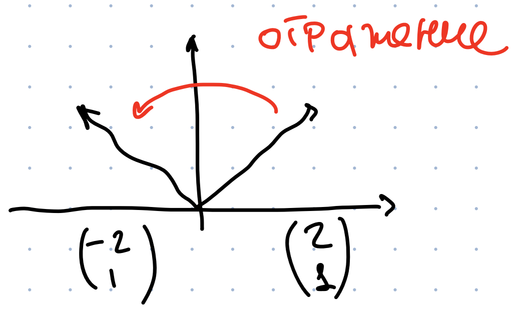
3.1 Спектральное разложение

$A: \mathbb{R}^n \to \mathbb{R}^n$

$A \cdot v = \lambda \cdot v \qquad \lambda \in \mathbb{R}$

λ — соб. зн.-я, v — соб. вектор

$\mathbb{R}^2$



$$\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$(A - \lambda \cdot I) v = 0$

$\det(A - \lambda \cdot I) = 0$

$\chi_A(\lambda)$ — хар. мн-н матрицы A

Утверждение 1

$\underset{[n \times n]}{A}$ если у неё $\exists$ n лин. нез. соб. векторов тогда $\boxed{A} = \begin{bmatrix} | & | & & | \\ p_1 & p_2 & \dots & p_n \\ | & | & & | \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} | & | & & | \\ p_1 & p_2 & \dots & p_n \\ | & | & & | \end{bmatrix}^{-1}$

$\qquad\qquad\qquad\qquad P \qquad\qquad D \qquad\qquad P^{-1}$

„диагонализация"

Утверждение 2

Спектральное разложение

A $[n \times n]$ $\qquad A^T = A$ симметричная

$$A = PDP^{-1} = PDP^T \qquad PP^T = I$$

$P^{-1} = P^T$ ортогональная

Если $A^T = A$ тогда

$\chi_A(\lambda) = 0$

имеет только действ. корни

$\lambda_1, \dots, \lambda_r \qquad n_1 + \dots + n_r = n$

$n_1, \dots, n_r$

Алгоритм:

$\det(A - \lambda_i I) = 0 \Rightarrow \lambda_1, \dots, \lambda_r$

$n_1, \dots, n_r$

$(A - \lambda_i I)x = 0$ $\qquad$ Для каждого λ_i

$\downarrow$ решение Гаусс LU

$x_1, \dots, x_{n_i}$

$\longrightarrow$ QR $\longrightarrow$ нормировка $\longrightarrow$ P

Если n действ. соб. векторов не $\exists$
неz.
Тогда можно спастись жордановой формой

$(1-\lambda)^2(2-\lambda)=0 \quad \lambda_1=1 \quad \lambda_2=1 \quad \lambda_3=2$

$$\begin{pmatrix} 1-\lambda & 0 & 0 \\ 1 & 1-\lambda & 0 \\ 3 & 2 & 2-\lambda \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 2 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 3 & 2 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$x_1=0$

$3x_1+2x_2+x_3=0$

$2x_2=-x_3$

$$\begin{pmatrix} 0 \\ -0.5 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0.5 \\ -1 \end{pmatrix}$$

$u_2 = v_2 - \alpha \cdot v_1$

$u_1 = \begin{pmatrix} 0 \\ 0.5 \\ -1 \end{pmatrix} \quad u_2 = \begin{pmatrix} 0 \\ 0.5 \\ -1 \end{pmatrix} - \frac{0.75}{1.25}\begin{pmatrix} 0 \\ -0.5 \\ 1 \end{pmatrix}$

$\alpha = \frac{0.75}{1.25}$

3.2. Сингулярное разложение (SVD)

любая

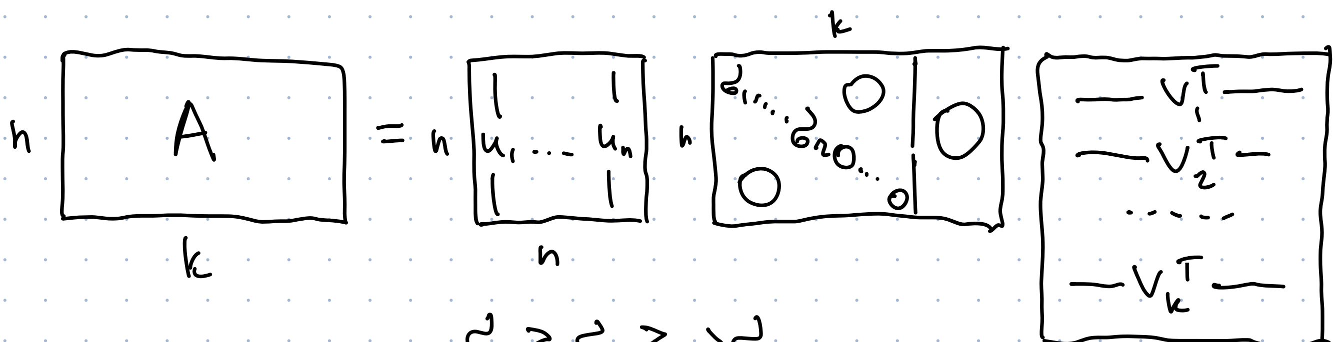
$$\underset{[n\times k]}{A} = \underset{[n\times n]}{U}\ \underset{[n\times k]}{\Sigma}\ \underset{[k\times k]}{V^T}$$

$V^T V = I$
$U^T U = I$ | ортогональные

Σ - диагональная

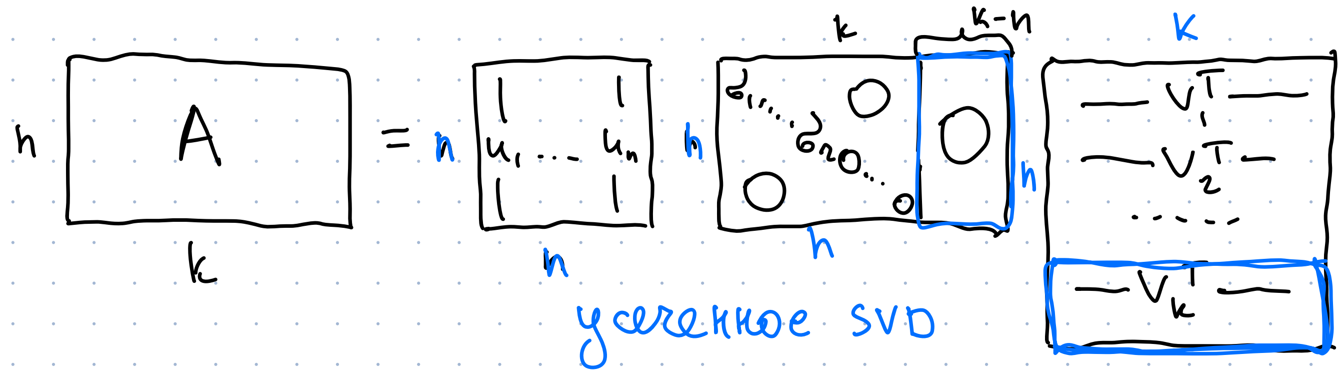
$$A = U \cdot \Sigma \cdot V^T$$

чёт делает | поворот | растяжение или сжатие | поворот

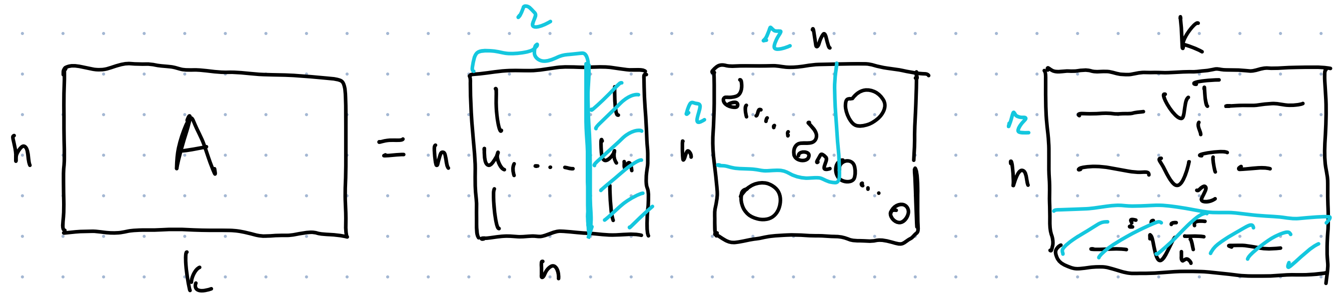


$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$

сингулярные числа матрицы A



усеченное SVD



компактный SVD

- → Сильно упрощаем вычисл.
- → Всегда ∃
- → Полезна ML
- → Показывает как устроен внутренний мир матрицы

$$A = U \Sigma V^T$$
$[2 \times 4]$

$A^T A$ $[4 \times 4]$ — симметричные полож. полуопред. матрицы

AA^T ✓ $[2 \times 2]$

$$A^TA = V\Sigma^T\underbrace{U^TU}_{I}\Sigma V^T = V\underbrace{\Sigma^T\Sigma}_{D}V^T = VDV^T$$

соб. вектора

соб. зн.

$$AA^T = U\underbrace{\Sigma\Sigma^T}_{D_1}U$$

U - соб. векторы знаем

Σ - $\sigma_1 = \sqrt{\lambda_1}$ .....

$(A \mid I) \rightsquigarrow (I \mid A^{-1})$

V - ?

$A = U\Sigma V^T$ $\quad AV = U\Sigma$ система ур.

Гаусс

$u_1 = \dfrac{Av_1}{\sigma_1}$, $u_2 = \dfrac{Av_2}{\sigma_2}$, ...

Алгоритм:

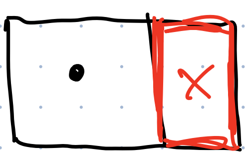
$A \qquad AA^T$

спектральное разложение для $AA^T \Rightarrow \Sigma, U$

Гаусс $\Rightarrow V'$

QR $\Rightarrow V$

+ нормировка

$V =$ 

находится неоднозначно

$\Rightarrow$ можно найти ● а потом дополнить до ортог. базиса

находится однозначно

Он медленный -
Комп. делает по-другому

Упражнение

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}}_{D} \underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}}_{U^T}$$

$$\chi_{AA^T}(\lambda) = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3$$

$\lambda_2 = 1 \quad \lambda_1 = 3 \qquad \sigma_2 = \sqrt{1} \quad \sigma_1 = \sqrt{3}$

$$\begin{pmatrix} 2-1 & 1 \\ 1 & 2-1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$v_1 + v_2 = 0$

$(1 \ \ -1) \quad (1/\sqrt{2} \ \ -1/\sqrt{2})$

$v_2 - v_1 = 0$

$(1, 1) \quad (1/\sqrt{2} \ \ 1/\sqrt{2})$

$$\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \cdot \frac{1}{\sqrt{3}} \qquad \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

$$V_1 + V_3 = \sqrt{3/2} \qquad V_1 + V_3 = \sqrt{1/2}$$

$$V_1 + V_2 = \sqrt{3/2} \qquad V_1 + V_2 = -\sqrt{1/2}$$

$V_1 = 0$ $\qquad$ $V_1 = 0$ в точку

$$\begin{array}{cc} V_1 & V_2 \\ \begin{pmatrix} 0 \\ \sqrt{3/2} \\ \sqrt{3/2} \end{pmatrix} & \begin{pmatrix} 0 \\ -\sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix} \end{array} \Rightarrow \text{ортогонализация } V_1 \perp V_2$$

$\|V_1\| = 3$

$\dfrac{\|V_1\|}{3}$

дополнить V_3 до ортогонального базиса $\begin{cases} \langle V_3, V_1 \rangle = 0 \\ \langle V_3, V_2 \rangle = 0 \end{cases}$

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \underset{[2\times2]}{\begin{pmatrix} 1/\sqrt2 & 1/\sqrt2 \\ 1/\sqrt2 & -1/\sqrt2 \end{pmatrix}} \underset{[2\times3]}{\begin{pmatrix} \sqrt3 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}} \underset{[3\times3]}{\begin{pmatrix} 0 & 0 & 1 \\ 1/\sqrt6 & -1/\sqrt2 & 0 \\ 1/\sqrt6 & 1/\sqrt2 & 0 \end{pmatrix}^T}$$

Примеры использования

① МНК

$$(X^T X)^{-1} X^T y$$

$$((U\Sigma V^T)^T (U\Sigma V^T))^{-1} (U\Sigma V^T)^T y$$

$$(V\Sigma^T U^T U \Sigma V^T)^{-1} V\Sigma^T U^T y$$

$$(V^T)^{-1} (\Sigma^T\Sigma)^{-1} V^{-1} V \Sigma^T U^T y$$

$$V (\Sigma^T\Sigma)^{-1} \Sigma^T U^T y$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Sigma^T\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \qquad (\Sigma^T\Sigma)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/9 \end{bmatrix}$$

сжимает

$$(X^TX)^{-1} = V(\Sigma^T\Sigma)^{-1}V^T$$

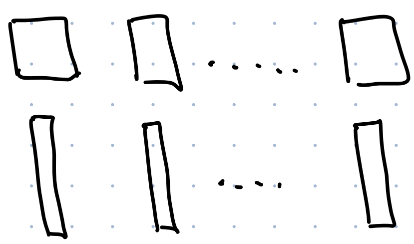
Остаются только умножения, нигде не надо искать обратные матрицы

$$X^TX = PDP^T$$

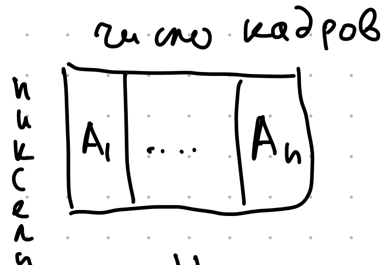
$$(X^TX)^{-1}\ (PDP^T)^{-1} = P^{-T}D^{-1}P^{-1} = PD^{-1}P^T$$

$$\sigma_i = \sqrt{\lambda_i} \qquad \frac{1}{\sigma_i^2} = \frac{1}{\lambda_i}$$

② Видеопоток



кадры



$$= U\Sigma V^T$$

Скорее всего, первые синг. числа отвечают за статическую часть

$\sigma_1\, u_1\, v_1^T$ – фон

③ PCA – метод главных компонент

$$\overset{k}{\underset{n}{\begin{bmatrix} x_1 & \ldots & x_k \end{bmatrix}}} \xrightarrow[se(x_i)]{x_i - \bar{x}_i} \overset{k}{\underset{n}{\begin{bmatrix} x'_1 & \ldots & x'_k \end{bmatrix}}}$$

$$\hat{\sigma}^2 = \overline{X^2} - \bar{X}^2 = \frac{1}{n-1}\sum_{j=1}^{n} x_{ji}^2 = \frac{1}{n-1} x_i^T x_i$$

$$\rho_{ik} = \widehat{cov}(x'_i, x'_k) = \frac{1}{n-1}\sum_{j=1}^{n} x'_{ij} x'_{kj} = \frac{1}{n-1} x'^T_i x'_k$$

$$\frac{X^T X}{n-1} = \begin{bmatrix} \hat{\sigma}_1^2 & \rho_{12} & \ldots \\ \rho_{12} & \hat{\sigma}_2^2 & \\ \vdots & & \ddots & \hat{\sigma}_k^2 \end{bmatrix}$$

$$X = \begin{bmatrix} | & | & | & \ldots & | \end{bmatrix}_{n \times D} \qquad \tilde{X}_{n \times d}$$

$$d < D$$

- PCA
- Низкоранг прибл.
- Тематич. модели W2V
- ALS
- Спектральная кластеризация

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

PCA на светую голову.

$$X = \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_k \\ \hline \end{array} \xrightarrow{x_j - \overline{x_j}} \begin{array}{|c|c|c|c|} \hline \tilde{x}_1 & \tilde{x}_2 & \dots & \tilde{x}_k \\ \hline \end{array}$$

$$\hat{Var}(x_i) = \overline{x^2} - \overline{x}^2 \qquad \mathbb{E}^2(x_i) \qquad \hat{Var}(\tilde{x}_i) = \overline{\tilde{x}_i^2} = \frac{1}{n-1}\sum \tilde{x}_i^2 = \frac{\tilde{x}_i^T \tilde{x}_i}{n-1}$$

$$\hat{Cov}(x_i, x_j) = \overline{x_i x_j} - \overline{x_i}\,\overline{x_j} \qquad \hat{Cov}(\tilde{x}_i, \tilde{x}_j) = \overline{\tilde{x}_i \cdot \tilde{x}_j} = \frac{\sum \tilde{x}_i \tilde{x}_j}{n-1} = \frac{\tilde{x}_i^T \tilde{x}_j}{n-1}$$

$$\mathbb{E}(x_i)\cdot\mathbb{E}(x_j)$$

$$\begin{pmatrix} \hat{Var}(\tilde{x}_1) & \dots & \hat{Cov}(\tilde{x}_1, \tilde{x}_n) \\ \hat{Cov}(\tilde{x}_2, \tilde{x}_1) & \ddots & \vdots \\ \vdots & & \\ \hat{Cov}(\tilde{x}_n, \tilde{x}_1) & \dots & \hat{Var}(\tilde{x}_n) \end{pmatrix} = \frac{1}{n-1}\begin{pmatrix} \tilde{x}_1^T\tilde{x}_1 & \tilde{x}_1^T\tilde{x}_2 & \dots & \tilde{x}_n^T\tilde{x}_n \\ \tilde{x}_2^T\tilde{x}_1 & \tilde{x}_2^T\tilde{x}_2 & \dots & \tilde{x}_2^T\tilde{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_n^T\tilde{x}_1 & \tilde{x}_n^T\tilde{x}_2 & \dots & \tilde{x}_n^T\tilde{x}_n \end{pmatrix} = \frac{\tilde{X}^T\tilde{X}}{n-1}$$

Буду считать, что все столбцы X центрированы $\tilde{X} = X$

$[k \times k]$

$\frac{X^T X}{n-1}$ – ковариационная матрица между признаками

$[n \times n]$

$\frac{X X^T}{k-1}$ – ков. мат. между объектами

PCA – хотим оставить в данных самое важное

Признаков очень много, хочется сжать их

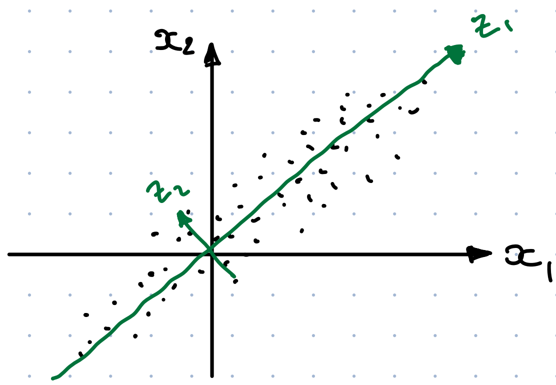
→ картинки (много пикселей одного цвета)

→ тексты (оч. много столбцов)

→ экспериментальные данные

① максимизация дисперсии

$$X = \underset{[n\times D]}{\begin{bmatrix} x_1 & x_2 & \dots & x_D \end{bmatrix}} \qquad \underset{[n\times d]}{X'} \qquad d \ll D$$



$$z_1 = u_{11} x_1 + u_{12} x_2 \qquad u_1 = \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix}$$

$$z_1 = X u_1$$

$$\begin{cases} Var(z_1) = z_1^T z_1 = u_1^T X^T X u_1 \to \max\limits_{u_1} \\ \|u_1\|^2 = u_1^T u_1 = 1 \end{cases}$$

$$[1\times D]\,[D\times n]\,[n\times D]\,[D\times 1]$$

$$\mathcal{L}(u, \lambda) = u_1^T X^T X u_1 - \lambda (u_1^T u_1 - 1)$$

$$\frac{\partial \mathcal{L}}{\partial u_1} = 2 X^T X u_1 - 2\lambda u_1 = 0$$

$$d(u_1^T X^T X u_1) = \underbrace{du_1^T X^T X u_1}_{[1\times 1]} + u_1^T X^T X du_1 = 2 u_1^T X^T X du_1$$

$$X^T X u_1 = \lambda u_1$$ u_1 – соб. вектор $X^T X$ соотв. λ_{max}

$$\begin{cases} Var(z_2) = u_2^T X^T X u_2 \to \max\limits_{u_2} \\ u_2^T u_2 = 1 \qquad \|u_2\| = 1 \\ u_2^T u_1 = 0 \qquad u_1 \perp u_2 \end{cases}$$

Лагранжиан

u_2 - соб. вектор $\lambda_{(2)}$

$$u_2^T X^T X u_2 - \lambda u_2^T u_2 - \mu u_1^T u_2$$

$$2 X^T X u_2 - 2\lambda u_2 - \mu u_1 = 0 \qquad \Big| \times u_2^T \text{ слева}$$

$$u_2^T X^T X u_2 - \lambda = 0 \qquad \Big| \times u_2 \text{ слева}$$

$$X^T X u_2 - \lambda u_2 = 0$$

$$X^T X u_2 = \lambda u_2$$

По аналогии получаем все компоненты

$$Var(x_1) + \dots + Var(x_D) = \frac{1}{n-1} tr(X^T X)$$

$$Var(z_1) + \dots + Var(z_D) = \frac{1}{n-1} tr(Z^T Z) =$$

$Z = X \cdot U$

$U^T U = I$

$$= \frac{1}{n-1} tr(U^T X^T X U) = \frac{1}{n-1} tr(X^T X)$$

$$X^T X u_k = \lambda_k u_k$$

$$\sum_{k=1}^{D} Var(z_k) = \frac{1}{n-1} \sum_{k=1}^{D} u_k^T X^T X u_k = \frac{1}{n-1} \sum_{k=1}^{D} \lambda_k$$

Хотим оставить d компонент $\qquad \dfrac{\lambda_1 + \dots \lambda_d}{\lambda_1 + \dots + \lambda_D}$

② PCA с т.з. SVD

X^TX λ

X^TX спектральное р-е

$\sigma_i = \sqrt{\lambda_i}$

$X = V \Sigma U^T$

$\underset{\text{новые набл.}}{Z} = \underset{\text{набл.}}{X} \underset{\text{веса}}{U} = V\Sigma$

$$Z^TZ = \Sigma^T V^T V \Sigma = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_D \end{bmatrix}$$

$$\underset{[n \times D]}{\begin{bmatrix} z_1 & z_2 & \dots & z_D \end{bmatrix}} = \underset{[n \times n]}{\begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}} \underset{[n \times D]}{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_D \\ 0 & 0 & 0 \end{bmatrix}}$$

③ PCA с т.з. минимизации ошибки приближения

$X_{n\times D}$ $\tilde{X}_{n\times D}$

$rk(X) = D$ $rk(\tilde{X}) = d \ll D$

Задача поиска малорангового приближения

$$\begin{cases} \|X - \tilde{X}\|_F^2 = \sum (x_{ij} - \tilde{x}_{ij})^2 = tr\left((X - \tilde{X})^T (X - \tilde{X})\right) \to \min\limits_{\tilde{X}} \\ \tilde{X} = ZU \\ U^TU = I \end{cases}$$

Th. Eckart-Young

4. Низкоранговые приближения

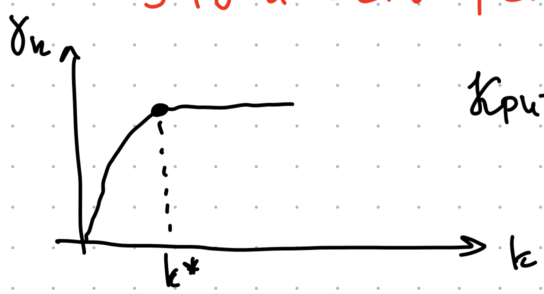
$$\begin{cases} \|A - B\|_F^2 \longrightarrow \min\limits_B \\ \operatorname{rk} B \leq k \end{cases}$$

$$A = \underbrace{\sigma_1 \cdot \boxed{u_1}\,\boxed{v_1^T} + \sigma_2 \boxed{u_2}\,\boxed{v_2^T} + \ldots}_{k} + \sigma_r \boxed{u_r}\,\boxed{v_r^T}$$

Это и есть решение

Какое k взять?

$$\gamma_k = \frac{\sigma_1^2 + \ldots + \sigma_k^2}{\sigma_1^2 + \ldots + \sigma_D^2}$$



Критерий крутого склона

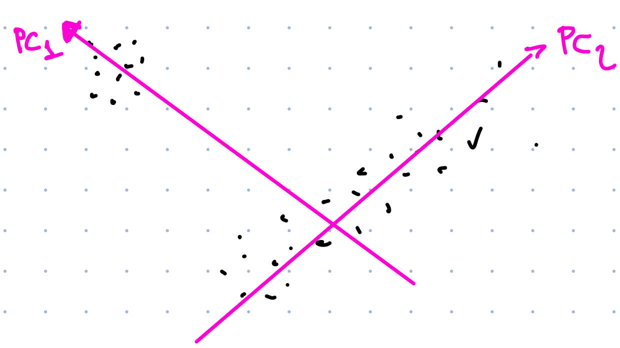
Скелетное р.-е.

$$\underset{[m\times n]}{A} = \underset{[m\times r]}{B} \times \underset{[r\times n]}{C}$$

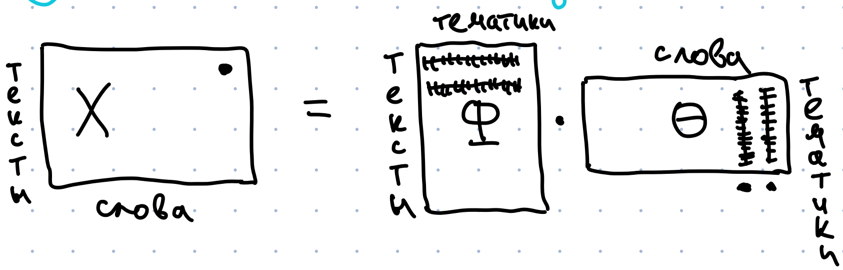
r - ранг A

r - маленькое. Тогда данные лежат очень компактно

- PCA $\mathbb{E}(x_i) = 0$
- SVD чувствителен к выбросам



(5.) Тематическое моделирование



∞ кол-во разложений
⇒ нужна регуляризация

$$P(w,d) = \sum_i \sum_j \underbrace{\varphi_{ik}}_{P(w|t)\cdot P(t|d)} \theta_{kj}$$

$$X = U \Sigma V^T \qquad \Phi \underbrace{S S^{-1}}_{I} \Theta$$

$$\Phi' \; \Theta'$$

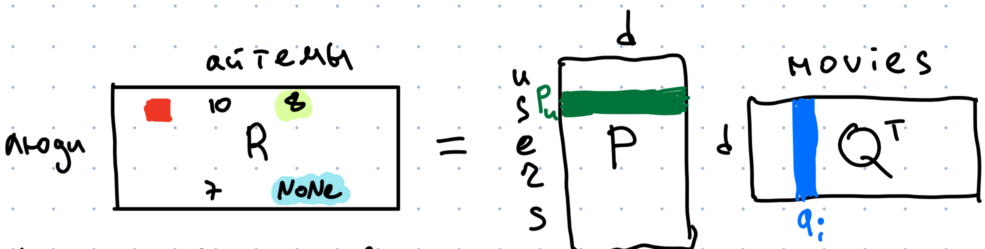
PLSA
LDA
ARTM

Развитие идеи: W2V
GloVe

Как использовать Φ — фичи

(6.) Рекомендации ALS

P, q — эмбеддинги



U — мн.-во юзеров
I — мн-во айтемов

SVD сделать не выйдет
⇒ Skeleton Decomposition
то же

$$\sum_{r_{ui} \text{ is not None}} \left(r_{ui} - \langle P_u, q_i \rangle\right)^2 \to \min_{P,Q}$$

прогноз: $\arg\max_j \langle P_{\text{Настя}}, q_j \rangle$

Сколько параметров надо выучить: $(|U| + |I|)\cdot d$

В модель можно добавить константы

$$r_{ui} = \langle p_u, q_i \rangle + \mu_u + \mu_i + \mu$$

SVD нельзя!! Как же обучаться?

$$Q = \sum_{r_{ui} \text{ is not None}} \big(r_{ui} - \overset{p_u^T q_i}{\langle p_u, q_i \rangle}\big)^2 \to \min_{Q,P}$$

$$\nabla_{p_u} Q = 2\sum_i \big(r_{ui} - \langle p_u, q_i \rangle\big)\cdot \underbrace{\begin{pmatrix} q_{i1} \\ \vdots \\ q_{id} \end{pmatrix}}_{q_i} = 0$$

$q_i \langle p_u, q_i \rangle$

$q_i q_i^T p_u$

$$\nabla_{q_i} Q = \dots$$

$$\sum_i r_{ui} \begin{pmatrix} q_{i1} \\ \vdots \\ q_{id} \end{pmatrix} - \sum_i q_i q_i^T p_u = 0$$

q_i $[d\times 1]$ $\quad p_u$ $[d\times 1]$

$$\forall u \quad p_u = \Big(\sum_i q_i q_i^T\Big)^{-1} \sum_i r_{ui} q_i$$

$$\forall i \quad q_i = \Big(\sum_u p_u p_u^T\Big)^{-1} \sum_u r_{ui} p_u$$

для текущего юзера и пересчет

ALS

0) init: p_u Q заполню рандомно
1) фиксирую P и делаю шаг по Q
2) фиксирую Q и делаю шаг по P

~10 шагов $\quad P \quad Q \quad$ click $\quad Q \equiv r_{ui}$

Наблюдение 1:

видос — 5 показов, 4 клика — $ctr = 4/5$

$\|q_i\| >> 0 \Rightarrow$ Переобучение

$\Downarrow$

Регуляризация

$R = PQ^T$ — решение задачи не единств.

$P S S^{-1} Q^T$

$P' Q'^T$ $\Rightarrow$ регуляризация

$$\sum_{r_{ui} \text{ is not None}} (r_{ui} - \langle p_u, p_i \rangle)^2 + \lambda \sum_u \|p_u\|^2 + \mu \sum_i \|q_i\|^2$$

Без обр. матрица: HALS

Hierarchical Alternating Least Squares

$$\frac{\partial Q}{\partial P_{ku}} = \sum_i 2(r_{ui} - \langle p_u, q_i \rangle) Q_{ki} = 0$$

$$\sum_i \Big(r_{ui} - \sum_{s \neq k} P_{su} Q_{si}\Big) Q_{ki} - P_{ku} \sum_i Q_{ki}^2 = 0$$

$$P_{ku} = \frac{\sum_i (r_{ui} - \sum_{s \neq k} P_{su} Q_{si})}{\sum_i Q_{ki}^2}$$

⑦ Спектральная кластеризация

X $\qquad$ $G(V, E)$ $\qquad$ ∀ способ (гиперпараметр) $\qquad$ пример

$\qquad\qquad W$

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\tau^2}\right)$$

Алгоритм:

$$D_{ii} = d_i = \sum_{j=1}^{n} w_{ij}$$

1) Строим G
2) Находим Лапласиан $L = D - W$

диагональная

3) Находим норм. соб. векторы для L

Берем те, которые соотв m наим. соб. значениям

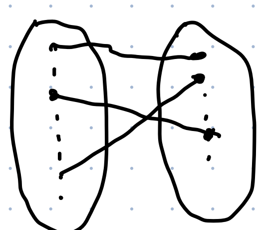
4) $U = \boxed{u_1 \,|\, \ldots \,|\, u_m}$

5) Обучаем на U алгоритм кластеризации K-means

Это то же самое что и Graph-cut подход

X $\quad$ W

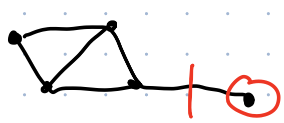
$$W(A,B) = \sum_{\substack{i \in A \\ j \in B}} w_{ij}$$



A $\qquad$ B

хочу выделить A и B убрав наим. кол-во инфы

$A \cap B = \varnothing$



$$\text{Ratio Cut}(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{W(A_i, \bar{A}_i)}{|A_i|} \to \min_{A_1, \dots, A_k}$$

k - число кластеров