SVD(奇异值分解)

1. 引言

$A \in R^{n\times d}$，秩是 $r \le \min(n,d)$

存在 ~~$\sigma_1, \sigma_2, \cdots$~~ $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$，$V_1, V_2, \cdots, V_r \in R^d$

$(V_i \perp V_j, \|V_i\|=1)$，$u_1, u_2, \cdots, u_r \in R^n$ $(u_i \perp u_j, \|u_i\|=1)$

使 $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ (A的奇异值分解)

记 $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$，$1 \le k \le r$

则 $\|A - A_k\|_2$，$\|A - A_k\|_F$ 最佳.

d维空间 R^d

$A = \begin{bmatrix} a^1 \\ \vdots \\ a^n \end{bmatrix}$ R^d中的n个d维向量.

$Sp(a^1, \cdots, a^n) = R(A)$ A的行空间.

$\subset R^d$，维数是 r

$Z(A) = \{x \in R^d, Ax = 0\}$，A的零空间. 维数：$d-r$

$R(A) \perp Z(A)$ $\quad R^d = R(A) \oplus Z(A)$

n维空间 R^n

$A = [a_1, a_2, \cdots, a_d]$，$a_j \in R^n$.

$C(A) \triangleq Sp(a_1, \cdots, a_d) \in R^n$

投影 (d维空间)

$x \in R^d$，$v \in R^d$，$\|v\|=1$

$P_v x \triangleq (x\cdot v)\cdot v \qquad \|P_v x\|^2 = (x\cdot v)^2$

$x \in R^d$，$S \subset R^d$，k维子空间.

$S = Sp(v_1, \cdots, v_k)$ $\quad (v_i \perp v_j, \|v_i\|=1)$

$P_S x = P_{v_1}x + \cdots + P_{v_k}x = (x\cdot v_1)v_1 + \cdots + (x\cdot v_k)\cdot v_k$

$\|P_S x\|^2 = (x\cdot v_1)^2 + \cdots + (x\cdot v_k)^2$

$x_1, \cdots, x_n \in R^d$，$S \subset R^d = Sp(v_1, \cdots, v_k)$

$P_S(x_1, \cdots, x_n) = (P_S x_1, \cdots, P_S x_n)$

$\|P_S(x_1,\cdots,x_n)\|^2 = \|P_S x_1\|^2 + \cdots + \|P_S x_n\|^2 = \|P_{v_1}X\|^2 + \cdots + \|P_{v_k}X\|^2$

$\|Ax\|^2 = \sum_{1}^{n} (a^i\cdot x)^2 \qquad \|Av\|^2 = \sum_{1}^{n} (a^i\cdot v)^2 = \|P_v A\|^2$

A的n个行向量在v上投影平方和

$\|Av_1\|^2 + \cdots + \|Av_k\|^2 = \|P_S A\|^2$

2. 矩阵奇异值和奇异向量

$\sigma_1(A) = \max_{\|v\|=1} \|Av\|$，$v_1 = \arg\max_{\|v\|=1} \|Av\|$

$\|Av_1\| = \sigma_1(A)$ $\quad (\|Av\| \le \sigma_1(A), \|v\| \le 1)$

$\sigma_2(A) = \max_{\substack{\|v\|=1 \\ v\perp v_1}} \|Av\|$，$v_2 = \arg\max_{\substack{\|v\|=1 \\ v\perp v_1}} \|Av\| \qquad \|Av_2\| = \sigma_2(A)$

$\cdots$

$\sigma_r(A) = \max_{\substack{\|v\|=1 \\ v\perp v_1, v_2, \cdots, v_{r-1}}} \|Av\|$，$v_r = \arg\max_{\substack{\|v\|=1 \\ v\perp v_1, \cdots, v_{r-1}}} \|Av\| \qquad \|Av_r\| = \sigma_r(A)$

SVD(2)

$u_i = \frac{1}{\sigma_i(A)} A v_i, \quad 1 \le i \le r$

定理 $S = Sp(w_1, \cdots, w_k)$ k维子空间 $\in R^d$，则

$\|P_S A\|^2 \le \|P_{(v_1,\cdots,v_k)} A\|^2$

证：对k用归纳法.

$k=1$ 正确.

若 $k=k$ 正确 $\Rightarrow k=k+1$ 也对.

$\cancel{Sp(v_1,\cdots,v_k)}$ $Sp(v_1,\cdots,v_k)^\perp$，d-k维

$Sp(w_1,\cdots,w_k,w_{k+1})$ k+1维

相交

存在 $w \in Sp(v_1,\cdots,v_k)^\perp \Rightarrow w \perp v_1,\cdots,v_k$

$\in Sp(w_1,\cdots,w_{k+1}) = Sp(w, w_1',\cdots,w_k')$

$\|P_{Sp(w,w_1',\cdots,w_k')} A\|^2 = \|P_w A\|^2 + \|P_{w_1'} A\|^2 + \cdots + \|P_{w_k'} A\|^2$

$= \|P_w A\|^2 + \|P_{Sp(w_1',\cdots,w_k')} A\|^2$

$\|P_{Sp(v_1,\cdots,v_{k+1})} A\|^2 = \|P_{v_1} A\|^2 + \cdots + \|P_{v_k} A\|^2 + \|P_{v_{k+1}} A\|^2$

$= \|P_{(v_1,\cdots,v_k)} A\|^2 + \|P_{v_{k+1}} A\|^2$

由 $\|P_{(v_1,\cdots,v_k)} A\|^2 \ge \|P_{Sp(w_1',\cdots,w_k')} A\|^2$，$\|P_{v_{k+1}} A\|^2 \ge \|P_w A\|^2$

$\Longrightarrow k=k+1$ 正确，

$A \in R^{n\times d}$

1. 预备知识

① 线性空间

$A = \begin{bmatrix} a^1 \\ \vdots \\ a^n \end{bmatrix}$，$a^i \in R^d$，$R(A) = Sp(a^1,\cdots,a^n) \perp Z(A) = \{x : Ax = 0\}$ （r，d-r）

~~A = [a, [a, ...,]]~~ $A = [a_1,\cdots,a_d]$，$a_j \in R^n$，$C(A) = Sp(a_1,\cdots,a_d)$

② 投影

2. A的奇异值和奇异向量

$v_i = \underset{\substack{\|x\|=1 \\ x \perp v_1,\cdots,v_{i-1}}}{\arg\max} \|Ax\|$，$\sigma_i(A) = \max_{\substack{\|x\|=1 \\ x\perp v_1,\cdots,v_{i-1}}} \|Ax\|$，$\sigma_i(A) = \|Av_i\|$

$1 \le i \le r$

$R(A) = Sp(v_1,\cdots,v_r)$，$v_{r+1},\cdots,v_d$，$Z(A) = Sp(v_{r+1},\cdots,v_d)$

定理 $Sp(v_1,\cdots,v_k)$，$1 \le k \le r$ 最优

3. SVD

① 定理： $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \equiv A_r$

引理： $A = B \Leftrightarrow \forall x,\ Ax = Bx$

$\Leftrightarrow \exists$ 基 $w_1,\cdots,w_d$，使 $Aw_i = Bw_i$，$1 \le i \le d$.

证(定理)： $v_1,\cdots,v_r; v_{r+1},\cdots v_d$

$A_r v_j = \left(\sum_{1}^{r} \sigma_i u_i v_i^T\right) v_j = \begin{cases} \sigma_j u_j, & 1 \le j \le r \quad = Av_j \\ 0, & j > r \quad = Av_j = 0 \end{cases}$

② $A_k = \sum_1^k \sigma_i u_i v_i^T$, $1 \le k \le r$

$D_k = A - A_k = \sum_{k+1}^r \sigma_i u_i v_i^T$, $1 \le k \le r$

$R(A_k) = Sp(v_1, \cdots, v_k)$, $Z(A_k) = Sp(v_{k+1}, \cdots, v_r, v_{r+1}, \cdots, v_d)$

$A_k v_j = \begin{cases} \sigma_j u_j = A v_j & 1 \le j \le k \\ 0 & j > k \end{cases}$

$R(D_k) = Sp(v_{k+1}, \cdots, v_r)$, $Z(D_k) = Sp(v_1, \cdots, v_k, v_{r+1}, \cdots, v_d)$

$\sigma_i(A_k) = \sigma_i(A)$, $1 \le i \le k$ $\quad \sigma_i(D_k) = \sigma_{k+i}(A)$ $1 \le i \le r-k$

$\sigma_i(A_k) = 0$, $k < i \le d$

③ 矩阵的模 $\|B\|_2 \triangleq \max_{\|x\|=1} \|Bx\|_2 = \sigma_1(B)$

定理：对一切秩小于等于 $k(\le r)$ 的矩阵 B，有

$\sigma_{k+1}(A) = \|A - A_k\|_2 \le \|A - B\|_2$

证明：$\|A-B\|_2 \ge \|(A-B)x\|$ $\quad (\|x\|=1)$

~~$Sp(v_1, \cdots, v_k, v_{k+1})$~~ $\underset{k+1}{Sp(v_1, \cdots, v_k, v_{k+1})}$, $\underset{d-k}{Z(B)}$

存在 $z \in Sp(v_1, \cdots, v_k, v_{k+1})$ $\quad \|z\| = 1$

$\in Z(B)$ $\quad z = \sum_1^{k+1} \alpha_i v_i$

~~$\|(A-B)z\|$~~ $\|(A-B)z\|^2 = \|Az\|^2$

$= \sum_1^{k+1} \alpha_i^2 (\sigma_i(A))^2 \ge \sigma_{k+1}(A)$

④ 矩阵的F模. ~~$\|B\|_F^2 \triangleq \sum_{i,j} b_{ij}^2 = \sum_{i=1}^n \|b^i\|^2 = \sum_{j=1}^d \|b_j\|^2$~~

$\|B\|_F^2 \triangleq \sum_i \sum_j b_{ij}^2 = \sum_{i=1}^n \|b^i\|^2 = \sum_{j=1}^d \|b_j\|^2 = \sum_1^r \sigma_i^2(B) = \sum_1^r \|B v_i(B)\|^2 = \|P_{(v_1, \cdots, v_r)} B\|^2$

证. $\sum_1^r \|B v_i(B)\|^2 = \sum_1^r \|P_{v_i} B\|^2 = \sum_{i=1}^n \|P_{(v_1, \cdots, v_r)} b^i\|^2 = \sum_{i=1}^n \|b^i\|^2$

定理 对一切秩小于等于 k 的矩阵 B，有：

$\|A - A_k\|_F \le \|A - B\|_F$

证：

$\|A\|_F^2 = \|P_S A\|^2 + \|A - S\|_F^2$, $\quad S = \begin{bmatrix} P_S a^1 \\ \vdots \\ P_S a^n \end{bmatrix}$

$\Uparrow$

$\|a^i\|^2 = \|P_S a^i\|^2 + \|a^i - P_S a^i\|^2$

$A_k = \begin{bmatrix} P_{(v_1, \cdots, v_k)} a^1 \\ \vdots \\ P_{(v_1, \cdots, v_k)} a^n \end{bmatrix}$

SVD (3)

1. 奇异值，奇异向量．特征值，特征向量

$A \in R^{n \times d}$，秩 $r \le \min(n, d)$

$\sigma_i, u_i, v_i, 1 \le i \le r \quad Av_i = \sigma_i u_i, A^T u_i = \sigma_i v_i$

$B \in R^{d \times d}$ 特征值，特征向量

对称：$B^T = B$

$B = A^T A \quad \lambda_i, v_i, 1 \le i \le r \quad (A^T A) v_i = \lambda_i v_i$

$\lambda_i = \sigma_i^2$

2. 特征向量的计算方法

① $(A^T A)^k$

$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T, \quad A^T = \sum_{j=1}^{r} \sigma_j v_j u_j^T$

~~$AA^T = (\sum_i \sigma_i u_i v_i^T)(\sum_j \sigma_j v_j u_j^T)$~~

$A^T A = (\sum_j \sigma_j v_j u_j^T)(\sum_i \sigma_i u_i v_i^T) = \sum_i \sigma_i^2 v_i v_i^T$

$(A^T A)^2 = \sum_i \sigma_i^4 v_i v_i^T, \quad (A^T A)^k = \sum_i \sigma_i^{2k} v_i v_i^T$

$= \sum_i \lambda_i^k v_i v_i^T$

如果 $\sigma_1 > \sigma_2$，$(A^T A)^k \approx \sigma_1^{2k} v_1 v_1^T$

② 设 $x \in R^d$，$x = \sum_{i=1}^{d} c_i v_i$

$(A^T A)^k \cdot x \approx (\sigma_1^{2k} v_1 v_1^T) \cdot (\sum_i^d c_i v_i) = \sigma_1^{2k} \cdot c_1 \cdot v_1$

$v_1 \approx \dfrac{(A^T A)^k \cdot x}{\|(A^T A)^k \cdot x\|}, \quad \sigma_1 = \|Av_1\|$

问题：x 选得不好，$c_1 = 0$.