SVD 奇异值分解. singular value decomposition

$A \in M_{m\times n}(\mathbb{R})$

(SVD) $A = QDP^T$.

$Q \in O(m)$ $P \in O(n)$

$$D = \left[\begin{array}{c} \begin{matrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m\end{matrix} \\ \hline 0 \end{array}\right] \quad m \geq n$$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$

A^TA 半正定, 对称, 正交对角化

eigenvectors $v_1, v_2 \cdots v_n \in \mathbb{R}^n$.

eigen values $\underbrace{\lambda_1, \lambda_2 \cdots \lambda_n}_{\geq 0} \in \mathbb{R}$ $\lambda_i = (\sigma_i)^2$

$P = (v_1\ v_2 \cdots v_n)$

$A^TA = (QDP^T)^T \cdot (QDP^T)$

$= P \cdot D^T \cdot Q^T \cdot Q D P^T$

$$= P \cdot \underline{(D^T D)}\, P^T .$$

$$(A^TA)\cdot(v_1 \cdots v_n) = \underline{(v_1 \cdots v_n)} \cdot$$

$$\underline{(A^TA)}\,\underline{P} = \underline{P}\cdot\begin{pmatrix} (\sigma_1)^2 & & & \\ & (\sigma_2)^2 & & \\ & & \ddots & \\ & & & (\sigma_n)^2 \end{pmatrix}$$

同样有 $Q = (w_1\ w_2 \cdots w_m)$

w_i 是 $A\cdot A^T$ 的特征向量.

特征值是 $(\sigma_i)^2$. 或者 0 $(i \geq n+1)$.

(事实. A $m\times n$. B $n\times m$, AB. BA. 除 0 以外的特征值相同.)

$(m \geq n)$ AB 比 BA 多 $m-n$ 个 0 特征值. (代数重数)

$\underline{m=n.\ |\lambda I - AB| = |\lambda I - BA|}$

$P = (v_1 \cdots v_n)$. $Q = (w_1 \cdots w_m)$

$$A = (w_1 \cdots w_m) \cdot \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix}$$

$$= \sum_{i=1}^{\min(m,n)} \sigma_i \cdot (w_i) \cdot v_i^T$$

A 来源　有 n 个变量，1，2，…，n

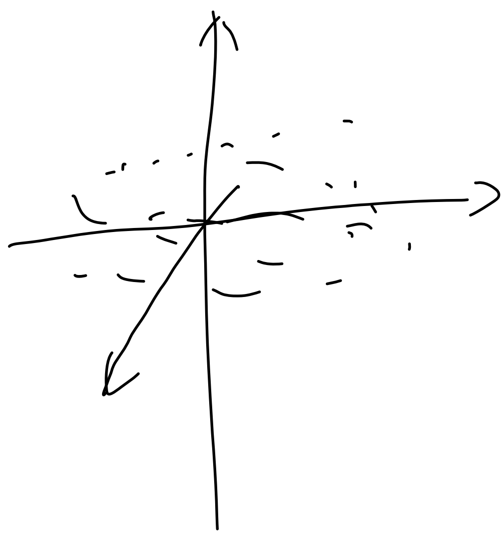
（每一种组分的浓度）

（温度，气压，风速）

$$A \begin{cases} (x_{11}, x_{12} \cdots x_{1n}) & \text{一次试验的结果} \\ & \text{(一位用户的数据)} \\ & \text{(一个点，一个时刻的天气)} \\ (x_{21}, x_{22}, \cdots x_{2n}) & \\ \vdots & \\ (x_{m1}, x_{m2}, \cdots x_{mn}) & \leftarrow m\text{次试验.} \end{cases}$$

A 行向量是每个点的数据

$$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}$$



n 维.

$$P^T = \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix}$$

标准正交基.

$$A = \underbrace{(QD)} \cdot P^T$$

P^T 是一组新的基 C

QD 的第 i 行 是 α_i^T 在 C 下的坐标.

直观下, $\underline{v_1^T, \cdots v_n^T}$ 方向的重要性

由 $\sigma_1 \geq \cdots \geq \sigma_n$ 来刻画.

低秩逼近.

定义: $A_r = Q \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \cdot P^T$

$\mathrm{rk}\, A_r \le r$

$= \sum_{i=1}^{r} \sigma_i w_i v_i^T$

$\underbrace{\left[\qquad\right]}_{r\text{列}} \cdot \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \cdot \underline{\left(\qquad\right)}$

Q的前r列

P^T的前r行

定理: $\|A - A_r\|_F \le \|A - B\|_F$, $\forall B,\ \mathrm{rk}\, B \le r$

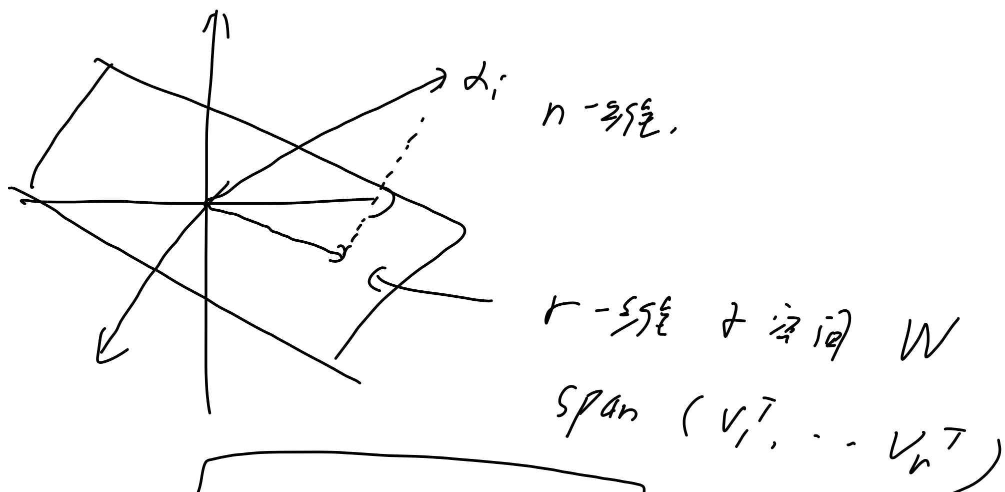
(Eckart-Young. Schmidt)

$\langle A, B \rangle_F = \mathrm{Tr}(A^T B)$

$A = (a_{ij}),\ B = (b_{ij}),\ \langle A, B \rangle_F$

$= \sum_i \sum_j a_{ij} b_{ij}$

低秩逼近的几何



$$A_r = \boxed{Q \cdot \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}} \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \\ 0 \end{pmatrix}$$

A_r 的第 i 行是 α_i 在 W 上的投影

$Q \cdot \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$ 第 i 行是该投影在 $v_1^T \cdots v_r^T$ 下的坐标.

Frobenius norm $\|\cdot\|_F$.

引理1. $C \in O(m)$. A, B $m \times n$.

$$\langle C\cdot A, C\cdot B\rangle_F = \langle A, B\rangle_F$$

1'. $C \in O(n)$. A, B $m \times n$

$$\langle AC, B\cdot C\rangle_F = \langle A, B\rangle_F$$

Min-Max property for singular value.

引理2. $\sigma_1 = \max\limits_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{|A\cdot v|}{|v|}$ $\leftarrow$ R quotient.

pf: $$\lambda_1(A^TA) = \boxed{\max\limits_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{\langle v, A^TAv\rangle}{\langle v, v\rangle}}$$

$$= \max\limits_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{\langle Av, Av\rangle}{\langle v, v\rangle}$$

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

引理3. $\boxed{\sigma_k(A - A_r) = \sigma_{k+r}(A)}$

$$A - A_r = Q \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{r+1} & & \\ & & & & \sigma_{r+2} & \\ & & & & & \ddots \\ & & & & & & 0 \\ & & & 0 & & & \end{pmatrix} P^T$$

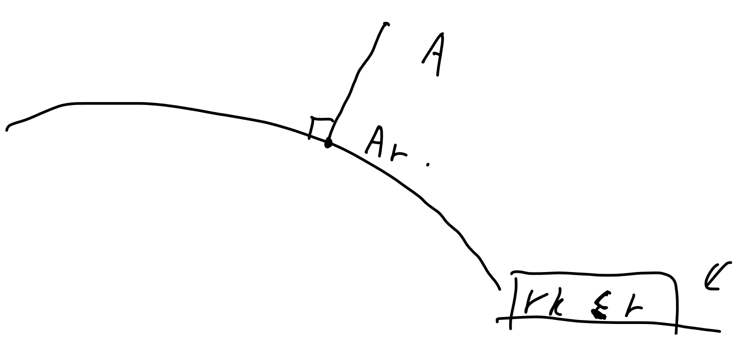
$A - A_r$ 的 SVD.

$$= \boxed{\sum_{k=1}^{\min(m,n)-r} \sigma_{r+k} \cdot u_{r+k} \cdot v_{r+k}^T}$$

推论: $\langle A, A \rangle_F = \|A\|_F$

$$= \sum_{i=1}^{\min(m,n)} (\sigma_i)^2$$

$$\|A - A_r\|_F = \sum_{i=r+1}^{\min(m,n)} (\sigma_i)^2 .$$



← 不是线性条件.

引理4: B $m\times n$. $\operatorname{rk} B \le r$

$$\sigma_1(A-B) \geq \sigma_{r+1}(A)$$

(对比 $\sigma_1(A-A_r) = \sigma_{r+1}(A)$)

Pf: 取 $v \in \underbrace{\operatorname{span}(v_1, \dots, v_{r+1})}_{\dim = r+1.} = W$

$\boxed{P \text{ 的前 } r+1 \text{ 列}}$

$\operatorname{rk} B \le r \Rightarrow$ 存在 $v \in W$. $v \neq 0$

$B \cdot v = 0$

$$\sigma_1(A-B) \geq \frac{|(A-B)\cdot v|}{|v|}$$

$$= \frac{|A\cdot v|}{|v|} = \sqrt{\frac{\langle v, A^T A v\rangle}{\langle v, v\rangle}}$$

$v = \sum_{i=1}^{r+1} a_i v_i$

$$= \sqrt{\frac{\sum_{i=1}^{r+1} (\lambda_i)\cdot a_i^2}{\sum_{i=1}^{r+1} a_i^2}} \geq \sqrt{\lambda_{r+1}} = \sigma_{r+1}$$

引理5: $\underbrace{\sigma_k(A-B)}_{}\geqslant \sigma_{k+r}(A)$

$\mathrm{rk}\, B \leq r$.

Pf: $\sigma_k(A-B) = \sigma_1\left((A-B)-(A-B)_{k-1}\right)$

$= \sigma_1\left(A - \left(B + (A-B)_{k-1}\right)\right)$

$\begin{cases} \mathrm{rk}\, B \leq r. \\ \mathrm{rk}\, (A-B)_{k-1} \leq k-1 \end{cases}$ $\quad \mathrm{rk}\left(B+(A-B)_{k-1}\right) \leq r+k-1$

(引理4) $\geqslant \sigma_{r+k}(A)$

定理的证明: $\|A-B\|_F^2$

$= \sum_{i=1} \left(\sigma_i(A-B)\right)^2$

$\geqslant \sum_{i=1} \left(\sigma_{i+r}(A-B)\right)^2$

$\geqslant \|A-A_r\|_F^2$

图像压缩. 2D 灰度图 像素 $m \times n$.

$$A = (a_{ij})$$

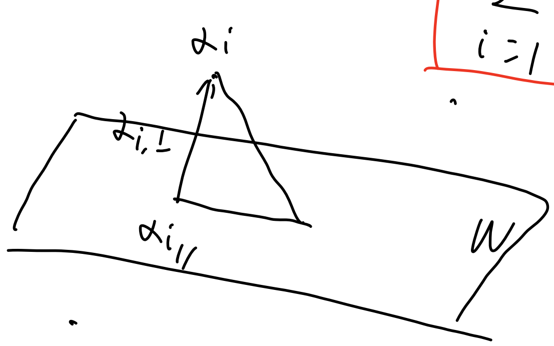
$\underline{r \ll \min(m,n)}$. $A_r = \underbrace{\begin{pmatrix} & \end{pmatrix}}_{r\text{列}} \begin{pmatrix} \sigma_1 & \cdots & \sigma_r \end{pmatrix} \cdot \begin{pmatrix} & \end{pmatrix}$ $\underline{r\text{行}}$

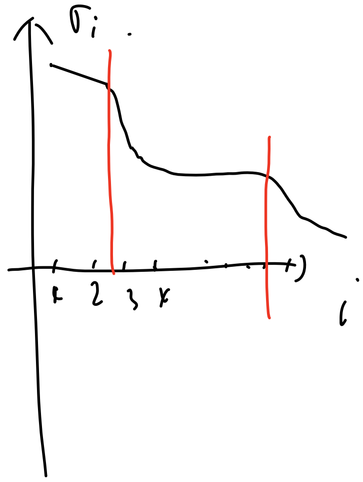
$m \gg n$ 　 m 次试验. 采样.

<u>n 维数据</u>

定理: $W_r = \mathrm{span}(v_1, \dots, v_r)$

W_r minimize $\displaystyle\sum_{i=1}^{m} (\mathrm{distance}(\alpha_i, W))^2$ among r-dim subspaces of $\mathbb{R}^n$





$\boxed{\text{人脸识别}}$ PCA

principal component analysis

(消除平均值的影响)

m次采样 取平均 $\frac{1}{m}\left(\sum_{i=1}^{m} x_i^T\right) = \mu$

$\bar{\mu} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \mu$.

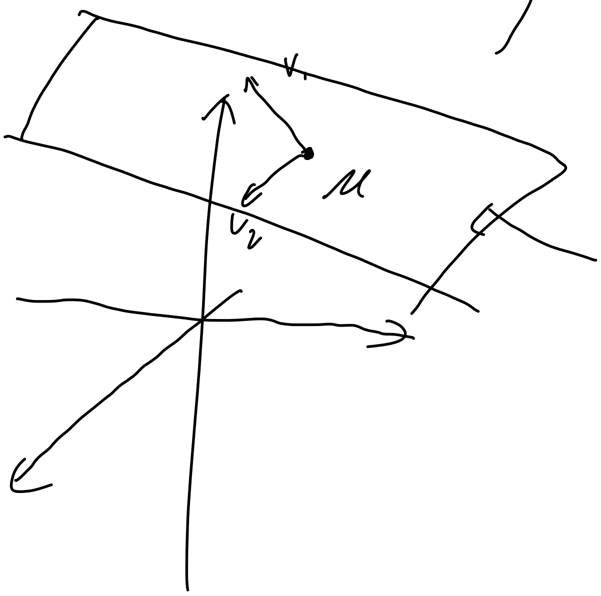
$B = (A - \bar{\mu})$. B的每一列平均值为0

$(A - \bar{\mu})_r + \bar{\mu}$ PCA

取 B 对应的 SVD

$B = Q\,D\,P^T$

$$B_r = Q \cdot \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix}$$



沿 μ 作平移的2维平面

逼近性质: $\underline{\mu + \text{span}(v_1, \dots v_r)}$

r维的平面(不一定经过原点)
仿射

与 $\alpha_1 \dots \alpha_m$ 的距离平方和最小的

(与 $\underline{SVD}$. dim $= r$ 逼近好
不比 $r+1$ 好)