SVD

$(m \geq n)$

$$A = QDP^T = \begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

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

$$A_k = Q \left[\begin{array}{c|c} \begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right] P^T = \left(\begin{array}{c|c} & 0 \end{array} \right) \left[\begin{array}{c|c} * & \\ \hline & \end{array}\right] \left[\begin{array}{c} \\ \hline 0 \end{array}\right]$$

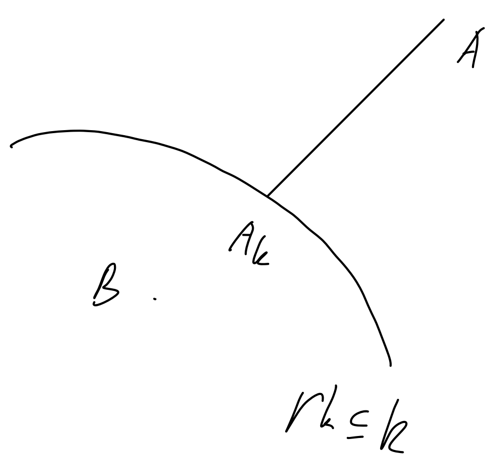
$$= \sum_{i=1}^{k} \sigma_i w_i v_i^T$$

定理 (Eckart-Young, Schmidt)

$$\| A_k - A \|_F \leq \| A - B \|_F \quad \forall B, \ \mathrm{rk}\, B \leq k$$

$$\| A - B \|_F = \sqrt{\sum_i \sum_j (a_{ij} - b_{ij})^2}$$

几何



证明中用到的重要性质（比一些技巧更重要）

① $\forall P \in O(n),\ Q \in O(m)$

$$\langle QAP^T, QBP^T \rangle_F = \langle A, B \rangle_F = Tr(A^T B)$$

② Min-Max for singular values

$$\sigma_1(A) = \max_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{|A \cdot v|_{\mathbb{R}^m}}{|v|_{\mathbb{R}}}$$

Pf: 回顾: min-max for eigenvalues.

$M = M^T \in M_n(\mathbb{R})$, $\lambda_1(M) \geq \lambda_2(M) \cdots \geq \lambda_n(M)$

$$\lambda_1(M) = \max_{\substack{v \neq 0 \\ v \in \mathbb{R}^n}} \frac{\langle v, Mv \rangle}{\langle v, v \rangle} \leftarrow \text{Rayleigh quotient}$$

(在对角化 M 的标准正交基 $v_1 \ldots v_n$ 下.

$v = \sum_{i=1}^{n} x_i v_i$, 则 Rayleigh quotient

$$= \frac{\sum_{i=1}^{n} \lambda_i x_i^2}{\sum_{i=1}^{n} x_i^2}$$
)

则 $\sigma_1(A) = \sqrt{\lambda_1(A^TA)}$

$$= \sqrt{\max_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{\langle v, A^TAv\rangle_{\mathbb{R}^n}}{\langle v, v\rangle_{\mathbb{R}^n}}} \quad = \frac{\langle Av, Av\rangle_{\mathbb{R}^m}}{\langle v, v\rangle_{\mathbb{R}^n}}$$

$$= \max_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{|Av|_{\mathbb{R}^m}}{|v|_{\mathbb{R}^n}}$$

$\square$

技巧：③（平移性质） $\sigma_i(A - A_k) = \sigma_{k+i}(A)$

$$A - A_k = \sum_{i=k+1}^{r} \sigma_i w_i v_i^T$$

④（平移不等式） $\forall B \in M_{m\times n}(\mathbb{R})$, $\operatorname{rk} B \leq k$

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

（对比 $\sigma_1(A - A_k) = \sigma_{k+1}(A)$）

Pf: 取 $W = \operatorname{span}_{\mathbb{R}}(v_1, \dots, v_{k+1}) \subset \mathbb{R}^n$.

$\dim W = k+1$.

$\mathrm{rk}\, B \le k$. $\dim \ker B \ge n-k$.

$\Rightarrow \ker B \cap W \ne 0$.

$\exists\, v \ne 0 \in \ker B \cap W$. $v = \sum_{i=1}^{k+1} a_i v_i$

$$\sigma_1(A-B) \ge \frac{|(A-B)\cdot v|_{\mathbb{R}^m}}{|v|_{\mathbb{R}^n}} = \frac{|A\cdot v|}{|v|}$$

$$= \frac{\sqrt{\sum_{i=1}^{k+1} \sigma_i^2 a_i^2}}{\sqrt{\sum_{i=1}^{k+1} a_i^2}} \ge \sigma_{k+1}.$$

⑤ (平移不等式加强) If $\mathrm{rk}\, B \le k$.

$$\sigma_l(A-B) \ge \sigma_{l+k}(A)$$

Pf:

$$\sigma_l(A-B) = \sigma_1\left((A-B) - (A-B)_{l-1}\right)$$

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

$\mathrm{rk}\, B \le k$. $\mathrm{rk}\,(A-B)_{l-1} \le l-1$.

$$\Rightarrow \operatorname{rk}\left(B+(A-B)_{l-1}\right) \leq k+l-1$$

则 $\sigma_1\left(A-\left(B+(A-B)_{l-1}\right)\right)$

(由 ④) $\geq \sigma_{k+l}(A)$

回到定理证明:

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

$$\geq \sum_{i=1}^{n}\left(\sigma_{i+k}(A)\right)^2$$

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

$$= \|A-A_k\|_F^2$$

□

相关应用

m 次试验，（采样） n 维数据 $(m >> n)$

$$A = \begin{pmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \vdots \\ \alpha_m^T \end{pmatrix} \qquad \alpha_i \in \mathbb{R}^n,$$

$\alpha = (x_1, \dots, x_n)$ $x_1 \dots x_n$ 之间可能存在相互依赖关系

或者 $\alpha_1 \dots \alpha_m$ 落在某个低维空间附近.

定理: $A = QDP^T$. $P = \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix}$

$W_k = \mathrm{span}_{\mathbb{R}}(v_1, \dots, v_k)$

则 W_k 使得如下值取最小.

$\sum_{i=1}^{m} (\mathrm{dist}(\alpha_i, W))^2$ 其中 W 是 $\mathbb{R}^n$ 的 $\dim = k$ 的子空间.

Pf: 对任一 W，取 W 的标准正交基

$u_1 \dots u_k$

则 $(\mathrm{dist}(\alpha, W))^2$

$$= \|\alpha - \mathrm{Proj}_W \alpha\|^2$$

$$= \|\alpha - \sum_{i=1}^{k} \langle \alpha, u_i \rangle u_i\|^2$$

$$= \| \alpha^T - (\langle \alpha, u_1 \rangle, \cdots \langle \alpha, u_k \rangle) \cdot \begin{pmatrix} u_1^T \\ \vdots \\ u_k^T \end{pmatrix} \|_F.$$

$$\sum_{i=1}^{m} \mathrm{dist}(\alpha_i, W)^2$$

$$= \| A - \underbrace{\begin{pmatrix} \langle \alpha_1, u_1 \rangle & \cdots & \langle \alpha_1, u_k \rangle \\ \vdots & & \\ \langle \alpha_n, u_1 \rangle & \cdots & \langle \alpha_n, u_k \rangle \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_k^T \end{pmatrix}}_{rk \le k} \|_F^2$$

另一方面: $\alpha = \sum_{i=1}^{n} \langle \alpha, v_i \rangle v_i$, $\alpha^T = (\langle \alpha, v_1 \rangle, \cdots \langle \alpha, v_n \rangle) P$

$A = QD \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$. QD 中第 i 行 $=$ $(\langle \alpha_i, v_1 \rangle \cdots \langle \alpha_i, v_n \rangle)$

所以 $QD \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \\ 0 \end{bmatrix} = \begin{pmatrix} (\alpha_1, v_1) & \cdots & (\alpha_1, v_k) \\ \vdots & & \\ (\alpha_m, v_1) & \cdots & (\alpha_m, v_k) \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_k^T \end{pmatrix}$

$= A_k$

由 Eckart-Young $\Rightarrow$ W_k 实现了

$\sum_{i=1}^{m} dist(\alpha_i, W)^2$ 的最小值.

消除均值的影响.

$\mu = \frac{1}{m}\left(\sum_{i=1}^{m} \alpha_i^T\right)$. $\bar{\mu} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \mu$.

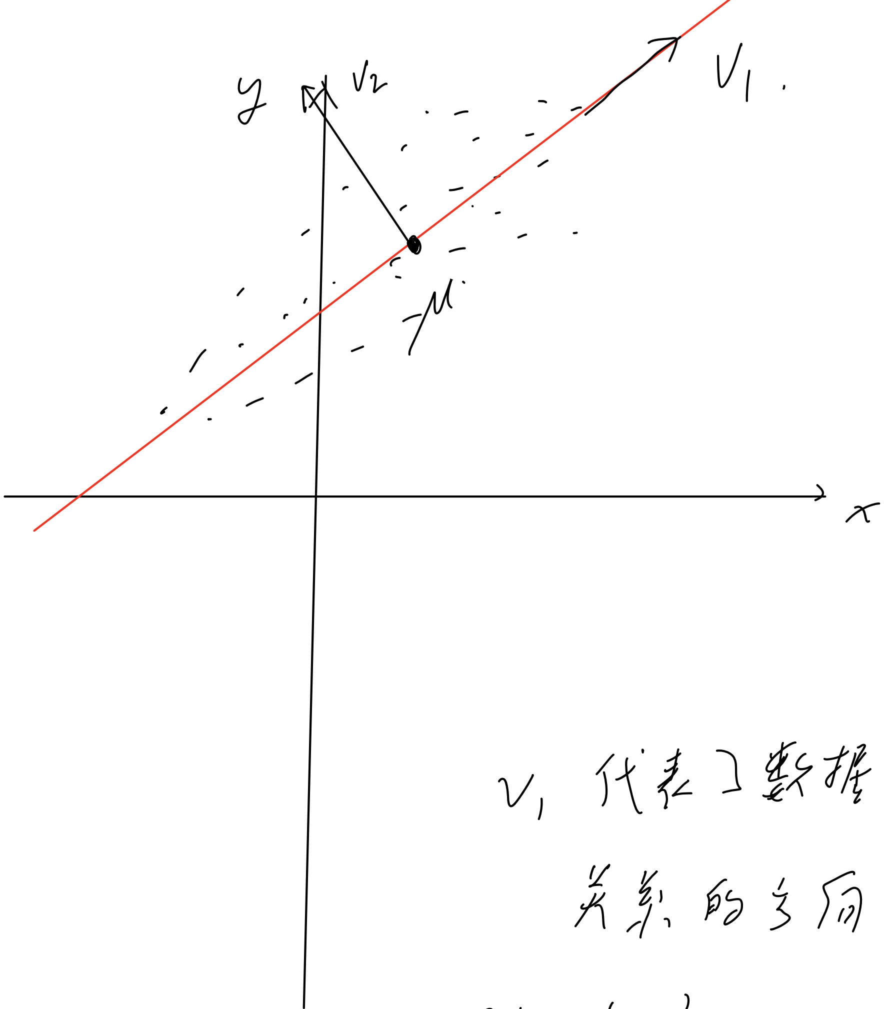
对 $B = A - \bar{\mu}$ 作 SVD

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

则 $\mu + span_{\mathbb{R}}(v_1 \cdots v_k)$ 是 k 维仿射平面中

与 $\alpha_1 \cdots \alpha_m$ 距离平方和最小的.

几何：



v_1 代表了数据的最主要的线性关系的方向.

v_2 次之，……

证明 (optional)

Step I:

对于固定 W,

取 $\beta \in \mathbb{R}^n$, minimize

$$\sum_{i=1}^{m} \left(dist(x_i, \beta + W) \right)^2$$

取 $\beta \perp W$,

$\text{dist}(\alpha, \beta + W)^2$

$= |\alpha - \text{Proj}_W \alpha - \beta|^2$

$\sum_{i=1}^{m} |(\alpha_i - \text{Proj}_W \alpha_i) - \beta|^2$

转化为 minimize $\sum_{i=1} |\gamma_i - \beta|^2$

则 $\beta = \frac{1}{m} \sum_{i=1}^{m} \gamma_i$. $\left(\sum_{i=1}^{m} (a_i - x)^2 \right.$ 在 $x = \frac{1}{m}\sum_{i=1}^{m} a_i$ 时 取最小值 $\left. \right)$

取最小.

即 $\beta = \frac{1}{m} \sum_{i=1} \alpha_i - \text{Proj}_W \frac{1}{m} \sum_{i=1}^{m} \alpha_i$

另一方面对 β 加上 W 中的向量不改变 $\beta+W$，所以

可取 $\beta=\frac{1}{m}\sum_{i=1}^{m}\alpha_i=\mu$.

$\sum_{i=1}^{m}\mathrm{dist}(\alpha_i,\beta+W)^2$ 的最小值在 $\beta=\mu$ 时取得.

回到 $\sum_{i=1}^{m}\mathrm{dist}(\alpha_i-\mu,W)^2$ 中对 W 的选取.

$$B=A-\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\mu.$$