SVD

$X_{n\times m}$ $\quad$ $\operatorname{rank}(X)=r$

$\rightarrow$ $\operatorname{rank}(X)=\operatorname{rank}(X^TX)=\operatorname{rank}(XX^T)$

$\rightarrow$ $X^TX,\ XX^T$ SDP $\quad$ Logo $\sigma(X^TX),\ \sigma(XX^T)\subset\mathbb{R}_0^+$

$$\begin{array}{ccccccccc} \lambda_1 & \geq & \lambda_2 & \geq & \lambda_3 & \geq & \cdots \geq & \lambda_r & > 0 \cdots \\ \downarrow & & \downarrow & & \downarrow & & & \downarrow & \\ v_1 & \perp & v_2 & \perp & v_3 & & \perp & v_r & \end{array} \qquad \text{SPG} \quad \|v_i\|=1$$

$X^TX = SDS^{-1}$

$D=\begin{bmatrix}\lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \ddots & & & \\ & & & \lambda_r & & \\ & & & & 0 & \\ & & & & & \ddots\ 0\end{bmatrix}$

$X^TXv_i=\lambda_i v_i \qquad v_i\neq 0$

$\sigma_i=\sqrt{\lambda_i}\in\mathbb{R}^+$

$u_i=\dfrac{1}{\sigma_i}Xv_i \qquad u_i\perp u_j \quad \|u_i\|=1$

mostra-se que

$$\begin{cases} Xv_1=\sigma_1u_1\\ Xv_2=\sigma_2u_2\\ \vdots\\ Xv_r=\sigma_ru_r\end{cases} \Rightarrow X\overbrace{\begin{bmatrix} | & & | & | & & | \\ v_1 & \cdots & v_r & v_{r+1} & \cdots & v_m \\ | & & | & | & & | \end{bmatrix}}^{V} =$$

$$=\underbrace{\begin{bmatrix} | & & | & \\ u_1 & \cdots & u_r & \cdots \\ | & & | & | \end{bmatrix}}_{=U}\underbrace{\begin{bmatrix}\sigma_1 & & & 0\\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r\end{bmatrix}}_{\Sigma}$$

as colunas de V têm $\|\cdot\|=1$ e são $\perp$ 2 a 2

$\Rightarrow V^{-1}=V^T \qquad (\ldots) \qquad U^{-1}=U^T$

$\sigma_i \equiv$ valores singulares de X

$\sigma_i = \sqrt{\lambda_i}$ $\quad \lambda_i$ é valor pr. de $X^T X$

que é valor pr. de $X X^T$

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

U, V são ortogonais ; $U^T = U^{-1}$; $V^{-1} = V^T$

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

$$X \approx \sum_{i=1}^{\tilde{r}} \sigma_i u_i v_i^T \qquad \tilde{r} < r$$

Aplicação no PCA

B matriz dos dados, centrados na média

$$B = \underbrace{\begin{bmatrix} | & | & | & | \end{bmatrix}}_{n \text{ indivíduos}} \Big\} \text{atributos}$$

$$S = \frac{1}{n} B B^T$$

$Y := \frac{1}{\sqrt{n}} B^T$ $\qquad$ $Y^T Y = S$

cada coluna de Y tem média 0

Aplicar SVD a $Y = U \Sigma V^{-1}$

As colunas de V são os vect. próp. de $Y^T Y$

P/ $X = B^T$, $X^T X = n Y^T Y$

Basta aplicar SVD a $X = B^T$

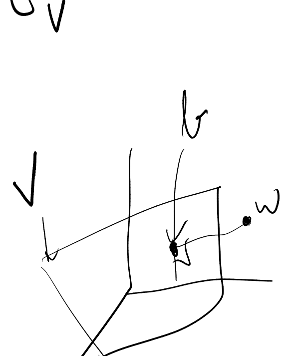
$X = U \tilde{\Sigma} V^{-1}$ onde as colunas de V

são os vect próp. de $X^T X$

Projecções!

$\text{proj}_V w$ é o único vector $b \in V$

t.q. $b = \underset{v \in V}{\arg\min} \|v - w\|$



$$\text{proj}_v w = \frac{\langle v, w \rangle v}{\|v\|^2} = \frac{(v^T w)}{v^T v} \cdot v$$

B é base ortogonal de V , $\mathcal{L}(B) = V$

$\| \{v_1, \ldots, v_n\}$ $\quad v_i \perp v_j$

SPG $\|v_i\| = 1$

$$\text{proj}_{\mathcal{L}(B)} b = \text{proj}_V b = \text{proj}_{v_1} b + \text{proj}_{v_2} b + \cdots + \text{proj}_{v_r} b$$

⚠ porque $v_i \perp v_j$

$$= (b^T v_1) \cdot v_1 + (b^T v_2) v_2 + \cdots + (b^T v_n) v_n$$

Sejam $v_1, \ldots, v_K$ compon. principais de B

ϕ é o novo input centrado na média do dataset

$$\text{proj}_{\mathcal{L}(v_1, \ldots, v_K)} \phi = ?$$

$$\text{proj}_{\mathcal{L}(v_1, \ldots, v_K)} \phi = \sum_{i=1}^{K} \text{proj}_{v_i} \phi = \sum_{i=1}^{K} \underbrace{(\phi^T v_i)}_{c_i} v_i$$

$$= \sum_{i=1}^{K} c_i v_i$$

$$c_i = \phi^T v_i \quad ; \quad C = \begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix}$$

$$\tilde{\phi} = \operatorname{proj}_{\mathcal{L}(v_1,\dots,v_k)} \phi = \begin{bmatrix} v_1 & \cdots & v_k \\ | & & | \end{bmatrix} \begin{bmatrix} \phi^T v_1 \\ \vdots \\ \phi^T v_k \end{bmatrix}$$

$$= V_k C$$

↳ pode ser obtido usando SVD

$$CS(V_k) = \mathcal{L}(v_1,\dots,v_k)$$

$$\tilde{x}_i = \operatorname{proj}_{\mathcal{L}(v_1,\dots,v_k)} x_i \qquad B = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix}$$

$$\operatorname{arg\,min}\ d(\tilde{\phi}, \tilde{x}_i)$$

$$d_n(\tilde{\phi}, \tilde{x}_i) = \sum_{j=1}^{k} \frac{1}{\lambda_j} \left((\tilde{\phi})_j - (\tilde{x}_i)_j \right)^2$$

Mahalanobis

$\lambda_1, \dots, \lambda_n$ valor. pr. BB^T

$k = ?$

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$$

Standardization min-max

$$\tilde{x}_i = a + \frac{(x_i - \min\{x_i\})(b-a)}{\max\{x_i\} - \min\{x_i\}}$$

standardization - z

$$\tilde{x}_i = \frac{x_i - \mu}{\sigma}$$ and $$\sigma = \sqrt{\frac{1}{N}\sum (x_i-\mu)^2}$$