Repaso de SVD

Dada $A \in \mathbb{K}^{m\times n}$, existen matrices:

- $U \in \mathbb{K}^{m\times m}$ unitaria $(U^* = U^{-1})$
- $\Sigma \in \mathbb{K}^{m\times n}$ diagonal
- $V \in \mathbb{K}^{n\times n}$ unitaria $(V^* = V^{-1})$

tales que $A = U \Sigma V^*$ y

$$\Sigma = \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & 0 \\ & & & \sigma_m & \end{array}\right) \quad \text{si } m \leq n \quad \left(A = \boxed{\quad\quad\quad\quad}\right)$$

ó

$$\Sigma = \left(\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{array}\right) \quad \text{si } m > n \quad \left(A = \boxed{\begin{array}{c} \\ \quad \\ \\ \end{array}}\right)$$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$ son los valores singulares.

$$A = U\Sigma V^* \iff \underbrace{A\cdot V}_{n\times m} = \underbrace{U\cdot\Sigma}_{n\times m}$$

Si $V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix}$, $U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{pmatrix}$

$$\Rightarrow A\cdot V = \begin{pmatrix} | & | & & | \\ Av_1 & Av_2 & & Av_m \\ | & | & & | \end{pmatrix}$$

Si $m \leq n$

$$U\cdot\Sigma = \begin{pmatrix} | & | & & | & \overbrace{0 \cdots 0}^{n-m \text{ ceros}} \\ \sigma_1 u_1 & \sigma_2 u_2 & & \sigma_m u_m & \\ | & | & & | & 0 \cdots 0 \end{pmatrix}$$

y

$m > n$

$$U\cdot\Sigma = \begin{pmatrix} | & & | \\ \sigma_1 u_1 & & \sigma_n u_n \\ | & & | \end{pmatrix}$$

En ambos casos $AV = U\cdot\Sigma$ dice que

$Av_i = \sigma_i u_i \quad i = 1\ldots, \min\{n,m\} = p$

si $m \leq n$ $\quad Av_i = \sigma_i \cdot u_i \quad i = 1\ldots m$

$\quad Av_i = 0 \quad i = m+1, \ldots n$

Si $m > n$ $\quad Av_i = \sigma_i u_i \quad i = 1 \dots n$

Av_i libre $\quad i = n+1, \dots, m$.

<u>OBS:</u> En ambos casos para los $\sigma_i = 0$ se tendrá $Av_i = 0$.

y en este caso vemos que $\sigma_i \cdot u_i = 0$ también $\Rightarrow$ esos u_i son libres.

¿Como construimos U, Σ, V?

1) Planteamos $A^*A \in \mathbb{K}^{n \times n}$ que es hermitiana

$\Rightarrow$ existe una BON de autovectores $v_1 \dots v_n$ y autovalores no negativos $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ (los ordenamos)

$$\Rightarrow V = \begin{pmatrix} \vdots & & \vdots \\ v_1 & \cdots & v_n \\ \vdots & & \vdots \end{pmatrix}, \quad \sigma_i = \sqrt{\lambda_i}$$

$$\Sigma = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{array}\right) \text{ ó } \Sigma = \left(\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{array}\right)$$

$$\sigma_i = \sqrt{\lambda_i}$$

2) Para $\sigma_i \neq 0$ $\quad u_i = \frac{1}{\sigma_i} A v_i$

$i = 1, \dots, s$

Para $\sigma_i = 0$ tenemos $A v_i = 0$

Para los $u_{s+1}, \dots, u_m$ que

faltan completamos a una BON

OBS: $A = U \Sigma V^* \iff A \cdot V = U \Sigma$

$A \in \mathbb{k}^{m \times n}$ $\quad V \in \mathbb{k}^{n \times n}$, $U \in \mathbb{k}^{m \times n}$

1) $\operatorname{Im} A = \{ A \cdot v : v \in \mathbb{k}^n \}$

$= \langle A w_1, \dots, A w_n \rangle$

para cualquier $\{w_1, \dots, w_n\}$ base de $\mathbb{k}^n$

Como $V = (V_1 | \dots | V_n)$ tiene una

BON en sus columnas $\Rightarrow$

$\operatorname{Im} A = \langle A v_1, \dots, A v_n \rangle$ $\quad p = \min\{n, m\}$

si asumimos $\quad = \langle \sigma_1 u_1, \dots, \sigma_p u_p \rangle$

$\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = 0$

$$= \langle \sigma_1 u_1, \dots, \sigma_r u_r \rangle$$

$$= \langle u_1, \dots, u_r \rangle$$

porque $\sigma_i \neq 0$ para $i = 1 \dots, r$

$$\Rightarrow \operatorname{Im} A = \langle u_1, \dots, u_r \rangle \text{ donde } r:$$

$$\sigma_1 \geq \dots \geq \underset{\substack{\neq \\ 0}}{\sigma_r} > \sigma_{r+1} = 0$$

2) $(\operatorname{Im} A)^\perp = \operatorname{Nu} A^*$

$$(\operatorname{Im} A)^\perp = \{ x \in \mathbb{K}^m : \langle x, w \rangle = 0 \ \forall w \in \operatorname{Im} A \}$$

$$= \{ x \in \mathbb{K}^m : \langle x, Av \rangle = 0 \ \forall v \in \mathbb{K}^n \}$$

$$= \{ x \in \mathbb{K}^m : \langle A^* x, v \rangle = 0 \ \forall v \in \mathbb{K}^n \}$$

donde usamos que $\langle x, Av \rangle = \overline{\langle Av, x \rangle}$

$$= \overline{\langle v, A^* x \rangle} = \langle A^* x, v \rangle$$

Además, $\langle A^* x, v \rangle = 0 \ \forall v \in \mathbb{K}^n$

si y solo si $\langle A^* x, e_i \rangle = 0 \; \forall e_i$ vector canónico. Esto nos dice que $(A^* x)_i = 0 \; \forall i = 1, \dots, n$ y entonces $A^* x = 0$

luego

$$(\operatorname{Im} A)^{\perp} = \{ x \in \mathbb{K}^n : A^* x = 0 \} = \operatorname{Nu} A^*$$

$\blacksquare$

Vimos la forma reducida:

Si $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = 0$

$$\Rightarrow A = U_r \Sigma_r V_r^* \in \mathbb{K}^{m \times n}$$

donde $V_r = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{pmatrix} \in \mathbb{K}^{r \times n}$

$$\Sigma_r = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \in \mathbb{K}^{r \times r}$$

$$U_r = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix} \in \mathbb{K}^{m \times r}$$

Por otro lado, si $k \leq r$

$$V_k = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{pmatrix} \quad \Sigma_k = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix}$$

$$U_k = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_k \\ | & & | \end{pmatrix}$$

$$A_k = V_k \Sigma_k U_k^*$$

$$= V_k (D_1 + \cdots + D_k) U_k^*$$

con $D_1 = \begin{pmatrix} \sigma_1 & & \\ & 0 \ddots & \\ & & 0 \end{pmatrix}$, $D_2 = \begin{pmatrix} 0 & & & \\ & \sigma_2 & & \\ & & 0 \ddots & \\ & & & 0 \end{pmatrix}$

$D_i = \begin{pmatrix} 0 & & & \\ & \ddots \sigma_i & & \\ & & 0 \ddots & \\ & & & 0 \end{pmatrix} \rightarrow$ lugar i

$$= V_k D_1 U_k^* + \cdots + V_k D_k U_k^*$$

$V_k D_i U_k^* =$

$$\begin{pmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{pmatrix} \begin{pmatrix} 0 & & & & \\ & \ddots 0 & & & \\ & & \sigma_i & & \\ & & & 0 \ddots & \\ & & & & 0 \end{pmatrix} \begin{pmatrix} \cdots & \overline{u_1} & \cdots \\ & \vdots & \\ \cdots & \overline{u_k} & \cdots \end{pmatrix}$$

$$= \begin{pmatrix} \vdots & & \vdots \\ v_1 & \cdots & v_k \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \cdots & \sigma_i & \bar{u}_i & \cdots \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \vdots & & \vdots \\ \sigma_i v_i \bar{u}_i^1 & \cdots & \sigma_i v_i \bar{u}_i^k \\ \vdots & & \vdots \end{pmatrix}$$

$$= \sigma_i \boxed{v_i} \cdot \boxed{u_i^*}$$

$$\Rightarrow A_k = \sigma_1 v_1 u_1^* + \cdots + \sigma_k v_k u_k^*$$

que tiene rango k.

Rango si $A = U \Sigma V^*$

$rg(A) = rg(\Sigma) = \#$ valores sing no nulos

Dem: $A \in \mathbb{K}^{m \times n}$

Vimos que $\operatorname{Im} A = \langle Av_1, \dots, Av_n \rangle$

$= \langle \sigma_1 u_1, \dots, \sigma_p u_p \rangle = \langle \sigma_1 u_1, \dots, \sigma_r u_r \rangle$

donde $\sigma_r > \sigma_{r+1} = 0$

$\Rightarrow rg(A) = \dim \operatorname{Im} A$

$= \dim \langle \sigma_1 u_1, \dots, \sigma_r u_r \rangle$

$= r = \#$ val sing no nulos.

Teorema:

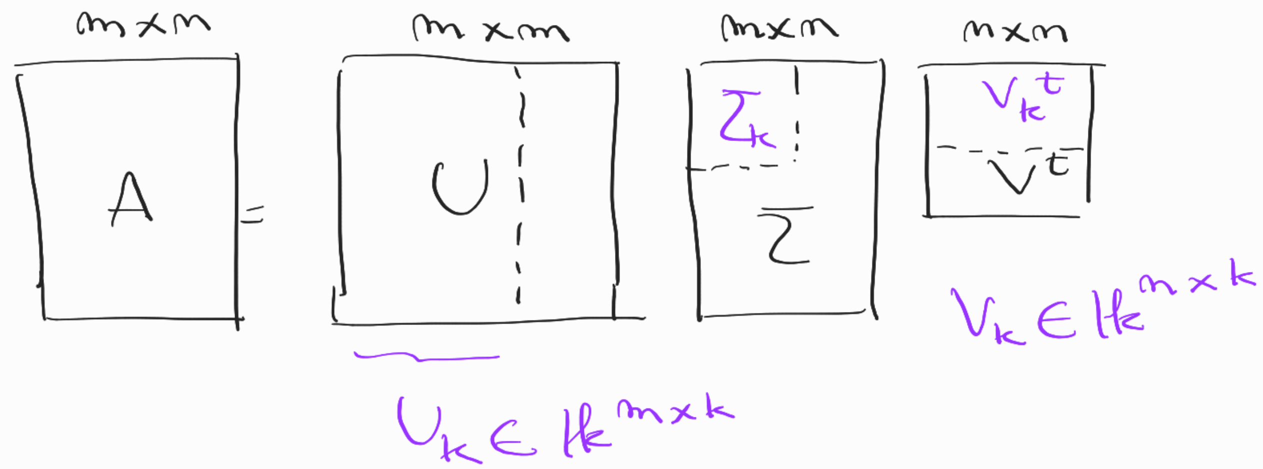
Vale que si $A \in \mathbb{K}^{m \times n}$ y $rg(A) = r \Rightarrow \forall\, k \leq r$ se tiene que A_k satisface

$$\|A - A_k\|_2 = \min_{B \in \mathbb{K}^{m \times n} :\ rg(B) = k} \|A - B\|_2 = \sigma_{k+1}$$

Es decir A_k es la "mejor" aproximación en norma 2 de matrices de la matriz A a una matriz de rango k.

Resumen

$m > n$



$A_k = U_k \Sigma_k V_k^*$ de rango $k < rg(A)$

<u>OBS:</u> $U_k^* U_k = Id_{k\times k}$, $U_k U_k^* \in \mathbb{K}^{m\times m}$

Recordar: Si S es un subesp. $\{w_1, \dots, w_k\}$ BON de $S \Rightarrow [P_S]_{EE} = \underbrace{(w_1|\dots|w_k)}_{W}\underbrace{(w_1|\dots|w_k)^*}_{W^*}$

es decir $P_S(x) = WW^*x$.

En este caso usaremos la notación P_{U_k} para la proyección en $S = \langle u_1, \dots, u_k\rangle$

es decir, $P_{U_k}(x) = U_k U_k^* x$

o análogamente $[P_{U_k}]_{EE} = U_k U_k^*$

Veamos que $A_k x = P_{U_k} A x$, es decir

veamos que $A_k = U_k U_k^* A$

$$U_k U_k^* A = U_k \underbrace{U_k^* U}_{} \Sigma V^*$$

$$U_k^* (U_k | u_{k+1} \cdots u_m)$$

$$= (Id \mid \underbrace{U_k^* [u_{k+1} \cdots u_m]}_{=0})$$

porque $u_1 \dots u_m$ es BON.

$\Rightarrow U_k U_k^* A$

$= U_k (Id | 0) \Sigma V^*$

$$= U_k (Id_k | 0) \left(\begin{array}{c|c} \Sigma_k & 0 \\ \hline 0 & \Sigma_{k+1 \dots m} \end{array} \right) \left(\begin{array}{c} V_k^* \\ \hline V_{k+1 \dots m}^* \end{array} \right) \to \text{el resto de los filas desde } \sigma_{k+1}$$

donde llamamos $\Sigma_{k+1\, m}$ al resto de la matriz diagonal de Σ desde el valor singular $\sigma_{k+1} \dots$

$\Rightarrow U_k U_k^* A$

$$= U_k \left(\Sigma_k \mid 0 \right) \left(\frac{V_k^*}{V_{k+1 \dots m}^*} \right)$$

$$= U_k \Sigma_k V_k^* = A_k \Rightarrow A_k = P_{U_k} A$$

Análogamente

$$V_k^* V_k = Id_{k \times k} \qquad V_k V_k^* = P_{V_k}$$

$$\Rightarrow A_k^* = P_{V_k} A^*$$

Es decir, la matriz A_k de rango k que es la que mejor aproxima a la matriz A en norma 2 viene dada por la proyección ortogonal de cada columna de A sobre el subespacio generado por las primeras k columnas de U.

Geométricamente esto nos dice que si tenemos muchos datos en $\mathbb{R}^3$

$$A = \begin{array}{c} \quad C_i(A) \\ \underbrace{\begin{bmatrix} \quad & \Big| & \quad \end{bmatrix}}_{3 \times n} \end{array}$$

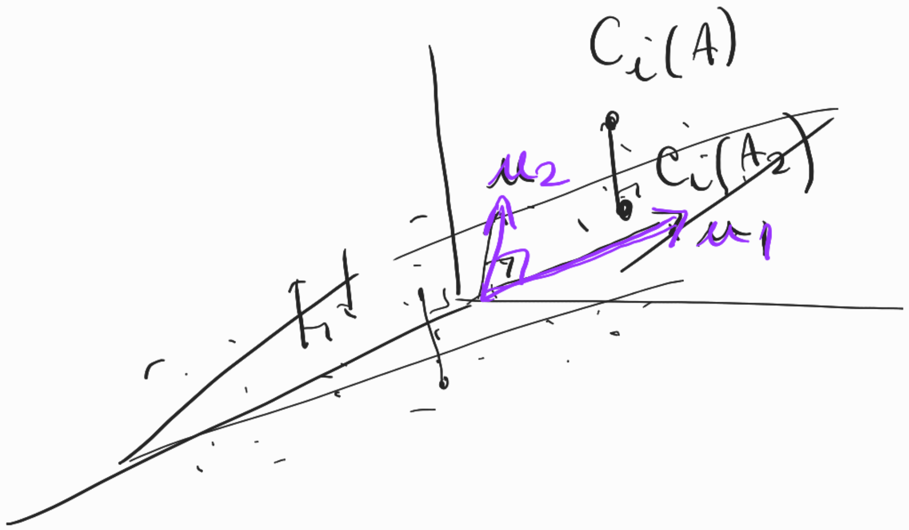
datos "centrados" con

$n >> 3 \quad rg(A) = 3$

A_2 la matriz de rango 2 que mejor aproxima a A en norma 2

es $A_2 = P_{\langle u_1, u_2 \rangle} A$

Es decir $\quad C_i(A_2) = P_{\langle u_1, u_2 \rangle} C_i(A)$



Valores singulares y norma 2

Vimos que:

<u>Teo</u>: $A \in \mathbb{K}^{m\times n} \Rightarrow \|A\|_2 = \sigma_1$

Además, si $A \in \mathbb{K}^{n\times n}$ es inversible

$\Rightarrow \|A^{-1}\|_2 = \dfrac{1}{\sigma_n}$ ya que

val sing de A^{-1} son $1/\sigma_i$, σ_i val

sing de A.

<u>Ejercicio</u>: Probar $A \in \mathbb{K}^{n\times n}$ es inversible

si y solo si $\sigma_n > 0$.

Luego, si A es inversible

$$\text{Cond}_2(A) = \sigma_1/\sigma_n.$$

<u>OBS</u>: $\|A\|_2 = \max\limits_{x\neq 0} \dfrac{\|Ax\|}{\|x\|} = \max\limits_{\|x\|=1} \|Ax\|$

Vimos que $\max\limits_{\|x\|=1} \|Ax\| = \sigma_1$

Además, $\|Ax\|_2 = \|U\Sigma V^* x\|_2 \underset{\substack{\uparrow \\ \text{Unitaria}}}{=} \|\Sigma \underbrace{V^* x}_{y}\|_2$

Si $y = V^* x \in \mathbb{K}^n$, $\Sigma \in \mathbb{K}^{m\times n}$

Si $\Sigma = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{array}\right)$ $\quad \Sigma \cdot y = \begin{pmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_m y_m \end{pmatrix}$ $\quad m \le n$

Si $\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{pmatrix}$ $\quad \Sigma y = \begin{pmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_n y_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $\quad m > n$

Si completamos $\sigma_{m+1} = \dots = \sigma_n = 0$ si $m \le n$ $\Rightarrow$

$$\|\Sigma y\|_2^2 = \sum_{i=1}^{n} (\sigma_i y_i)^2 = \sum_{i=1}^{n} \sigma_i^2 y_i^2 \geq \sigma_n^2 (y_1^2 + \dots + y_n^2)$$

$$\Rightarrow \|\Sigma y\|_2 \geq \sigma_n \|y\|_2$$

$$\Rightarrow \|\Sigma V^* x\|_2 \geq \sigma_n \|V^* x\|_2 = \sigma_n \|x\|_2$$

$$\Rightarrow \|Ax\|_2 \geq \sigma_n \|x\|_2 \Rightarrow \|Ax\|_2 \geq \sigma_n$$

si $\|x\|_2 = 1$

y como $\|A v_n\|_2 = \sigma_n$ tenemos

$$\min_{\|x\|_2=1} \|Ax\|_2 = \sigma_n.$$ (Que será 0 si $m < n$)

Pseudoinversa (de Moore-Penrose)

Da la SVD de $A = U \Sigma V^*$ construimos la pseudo inversa de A

$$A^\dagger = V \Sigma^\dagger U^* \quad \text{donde } \Sigma^\dagger$$

se construye trasponiendo Σ y cambiando cada $\sigma_i \neq 0$ por $\frac{1}{\sigma_i}$

Proposición:

1) $A A^\dagger A = A$ y $A^\dagger A A^\dagger = A^\dagger$

(pero en general no se cumple que $A A^\dagger = Id$ ó $A^\dagger A = Id$)

2) Si $A \in \mathbb{K}^{n \times n}$ es invertible $\Rightarrow A^\dagger = A^{-1}$

Dem:

1) $U\Sigma \underbrace{V^* V}_{Id} \Sigma^{\dagger} \underbrace{U^* U}_{Id} \Sigma V^*$

$$= U \Sigma \Sigma^{\dagger} \Sigma V^*$$

Si $\Sigma = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right)$ $\sigma_r \neq 0$

$$\Sigma^{\dagger} = \left(\begin{array}{ccc|c} \sigma_1^{-1} & & & \\ & \ddots & & 0 \\ & & \sigma_r^{-1} & \\ \hline & 0 & & 0 \end{array}\right)$$

$$\Rightarrow \Sigma \Sigma^{\dagger} \Sigma = \Sigma \quad ✓$$

Ej: Demostrar lo que falta

2) A inversible $\Leftrightarrow \sigma_n > 0$ .....

<u>teo</u>: $Ax = b$ $A \in \mathbb{K}^{m \times n}$. Si existen soluciones $\Rightarrow z = A^{\dagger} b$ es una sol y si hay infinitas, z es la de menor norma 2.

Teo: $Ax=b$ tiene sol sii $AA^{+}b=b$

Dem:

$\Rightarrow)$ sup $\exists\, v$: $Av=b$

$$\Rightarrow b = Av = AA^{+}Av = AA^{+}b$$

$\Leftarrow)$ $A\underbrace{A^{+}b}_{x}=b \Rightarrow \exists\, x$: $Ax=b$.

Teo: Dado $A\in \mathbb{K}^{m\times n}$, $b\in \mathbb{K}^{m}$

si $AA^{+}b=b$, x es sol de $Ax=b$

sii $x = A^{+}b+(I-A^{+}A)y$ con

$y\in \mathbb{K}^{n}$

Dem:

$\Leftarrow)$ $A\left(A^{+}b+(I-A^{+}A)y\right)$

$$= AA^{+}b+Ay-\underbrace{AA^{+}Ay}_{Ay}$$

$$= AA^{\dagger}b = b$$

$\Rightarrow$) Dado z: $Az = b$

$$z = A^{\dagger}\underbrace{Az}_{b} + \underbrace{(z - A^{\dagger}Az)}_{(I - A^{\dagger}A)\underbrace{z}_{y}}$$

<u>OBS:</u> pueden estar repitiéndose soluciones.

<u>Ejercicio:</u> $A \in \mathbb{K}^{m \times n}$

Si $m > n$ y $rg(A) = n \Rightarrow$

$$A^{\dagger} = (A^{*}A)^{-1}A^{*} \quad y \quad A^{\dagger} \cdot A = Id$$

Si $m < n$ y $rg(A) = m \Rightarrow$

$$A^{\dagger} = A^{*}(AA^{*})^{-1} \quad y \quad A \cdot A^{\dagger} = Id$$

Aplicaciones

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times m}\ \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{pmatrix}}_{m\times n}\ \underbrace{V^*}_{n\times n}$$

Si $rg(A) = r < n \Rightarrow \sigma_{r+1} = \cdots = \sigma_n = 0.$

$$\Rightarrow A = A_r = \underbrace{\begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix}}_{m\times r}\underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}}_{r\times r}\underbrace{\begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_r & - \end{pmatrix}}_{r\times n}$$

OBS: $$\underbrace{\begin{pmatrix} --- & v_1 & --- \\ & & \\ --- & v_r & --- \end{pmatrix}}_{r\times n}\underbrace{\begin{pmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{pmatrix}}_{n\times n} = Id_r$$

$$\Rightarrow A\cdot\begin{pmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ u_1 & & u_r \\ | & & | \end{pmatrix}\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

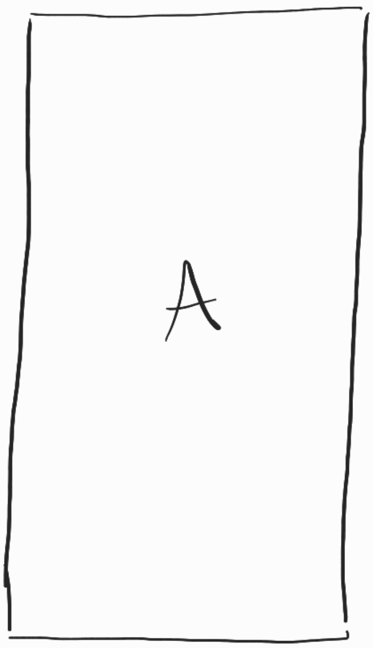
El resto $v_{r+1} \cdots v_n$ están en $Nu(A)$.

Aplicaciones:

- Reducción de dimensionalidad.

n características (altura, peso, etc)

indiv m



Puede haber características redundantes

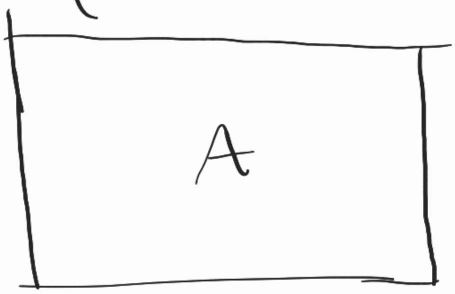
$rg(A) = r$

si $k < r$

σ_{k+1} nos dice cuánta información perdemos si usamos A_k en lugar de $A = A_k$.

- Imágenes (en color, grises)

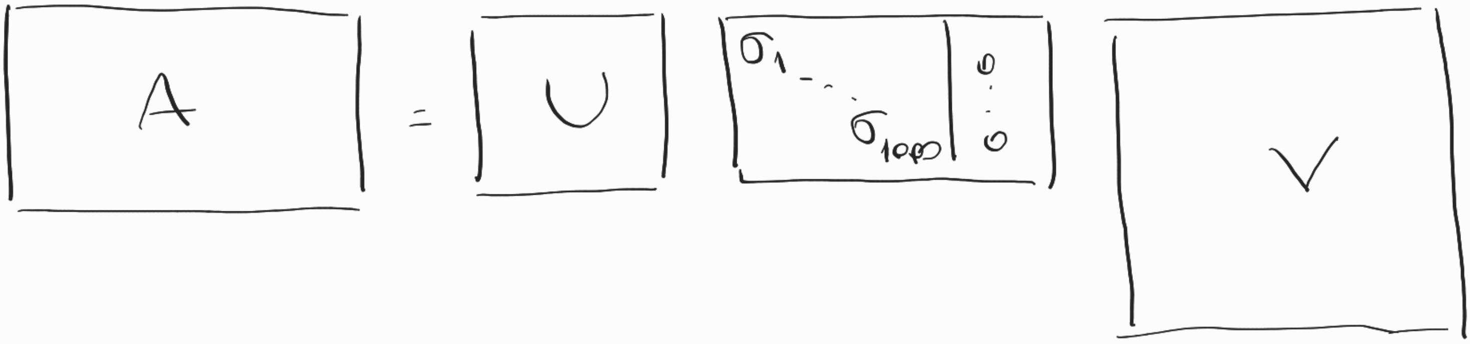
Grises 1000



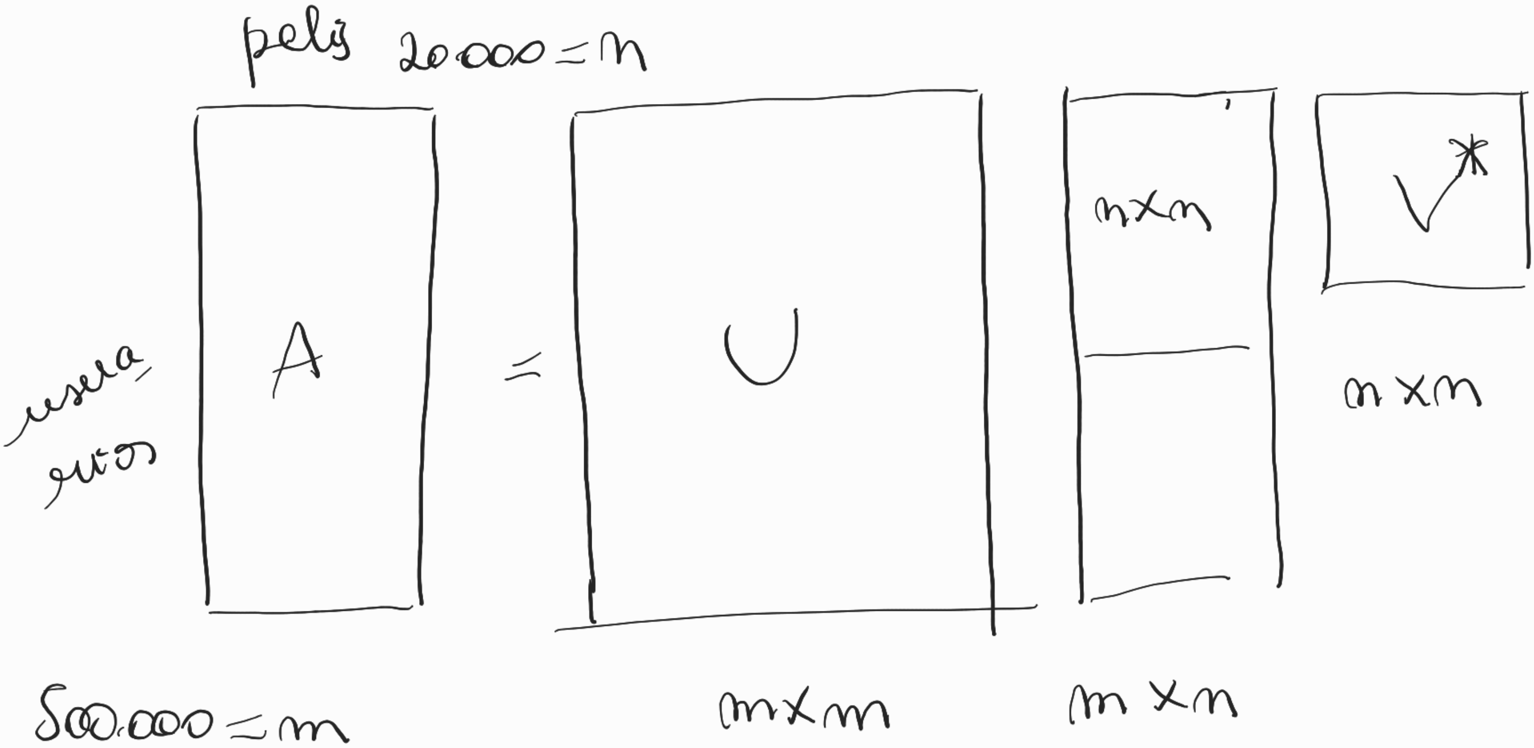
2000

pixeles 1000×2000

¿Cuánto puedo comprimir?



- Algoritmos de recomendación



Uso reducida: $r = rg(A)$

$$A = A_r = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_r & - \end{pmatrix}$$

$m \times r$ $\qquad\qquad\qquad\qquad$ $r \times n$

$$A = \underset{\text{usuarios}}{\overset{\text{"características"}}{\boxed{U_r}}} \; \boxed{\begin{matrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{matrix}} \; \overset{\text{pelis}}{\boxed{V_r^*}}\Big\}\text{"características"}$$

$U_r \downarrow$ caract "más relevante"

$\sigma_1, \dots, \sigma_r \downarrow$ importancia de c/ característica

Las "características" son desconocidas.

Cada

$$a_{ij} = \underbrace{q_i^t}_{\text{concentra características de usuario } i} \cdot \underbrace{p_j}_{\text{concentra relación entre caract y peli } j}$$

$q_i \to$ fila i de $\sqrt{\Sigma}\, U_r = Q_r$

$p_j \to$ col j de $\sqrt{\Sigma}\, V_r^* = P_r$

Se busca

$$\min_{p,q} \sum_{(i,j)\in\Omega} \big(r_{ij} - \overbrace{q_i^t p_j}^{a_{ij} \text{ de la q' queremos}}\big)^2 + \lambda \|Q_r\|_F^2 + \lambda \|P_r\|_F^2$$

$\Omega = \{(i,j) : r_{ij}$ es conocido$\}$

Conjunto de entre-namiento.

- Principal Components Analysis

Idea es dar un sistema de coordenadas jerárquico basado en los datos, características

$$A = \begin{pmatrix} - & a_1 & - & - & - \\ - & a_2 & - & - & - \\ & & & & \end{pmatrix} \to \text{persona}$$

$\in \mathbb{K}^{m \times n}$

Cada a_i es un conjunto de medidas de un mismo experimento

Por ej datos demográficos: altura, peso, género, etc.

y cada fila sería un individuo

$(m >> n)$

Queremos determinar las características dominantes de los datos.

1. Calculamos la media (promedio) vectorial

$$\bar{a} = \frac{1}{n} \sum_{j=1}^{n} a_j$$

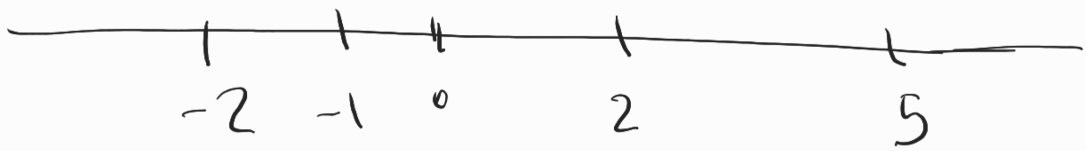
En cada lugar $\bar{a}_i$ está el promedio de c/ característica y definimos

$$\bar{A} = \begin{pmatrix} \cdots & \bar{a} & \cdots \\ \cdots & \bar{a} & \cdots \\ & \vdots & \\ \cdots & \bar{a} & \cdots \end{pmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot [\bar{a}]$$

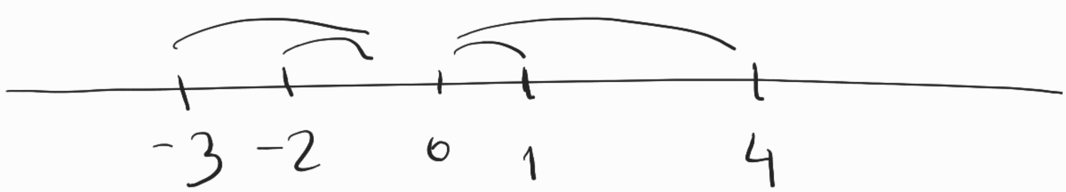
$\Rightarrow$ $B = A - \bar{A}$ es una matriz que tiene los datos "centrados".

<u>Ej:</u> en $\mathbb{R}$

$a_1 = -2$, $a_2 = -1$
$a_3 = 2$ $a_4 = 5$



$$\bar{a} = \frac{-2-1+2+5}{4} = \frac{4}{4} = 1$$



var

$$\frac{1}{4}\left((-3)^2+(-2)^2+1^2+4^2\right)$$

2. Calculamos autovalores y autovectores de $B^t \cdot B$

(es decir, calculamos los $\sigma_i = d_i^2$ valores ppales y los vectores V de la SVD de B)

$$B = U \Sigma V^t$$

nuevas direcciones

$$T = B \cdot V = U \Sigma$$

componentes ppales

1^a comp a $B \cdot v_1 = \sigma_1 u_1$

2^{da} comp $Bv_2 = \sigma_2 u_2$ . . . .

$B v_1 = v_1^1 C_1(B) + \dots + v_1^n C_n(B)$ con $v_1 = (v_1^1, \dots, v_1^n)$

Es decir $\sigma_1 u_1$ es una comb. lineal de col de la matriz de datos B.

$B^t B$ = matriz de covarianza

$$B^t B = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

$$B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \begin{array}{l} \to \text{individuo } 1 \\ \\ \to \text{individuo } n \end{array}$$

$\uparrow \uparrow \uparrow$ m características

peso altura

Datos en $\mathbb{R}^n$

$$B^t \cdot B = \begin{pmatrix} b_1^1 & \cdots & b_n^1 \\ \\ b_1^m & \cdots & b_n^m \end{pmatrix} \begin{pmatrix} b_1^1 & & b_1^m \\ \vdots & \cdots & \\ b_n^1 & & b_n^m \end{pmatrix}$$

$\downarrow$ característi de indiv 1

$\downarrow$ indiv n

$\downarrow$ caract 1 de c/indiv

$$\langle b_1 , b_1 \rangle = \|b_1\|^2$$

$$\langle b_1 , b_2 \rangle = \|b_1\| \, \|b_2\| \cos\theta$$

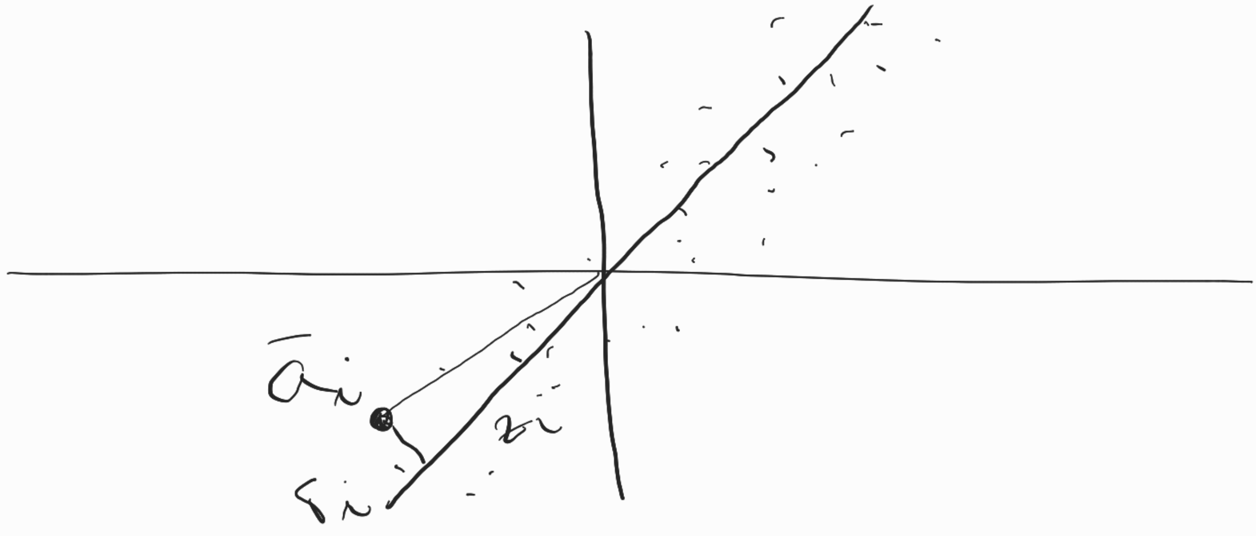
será mayor si $\theta = 0$

y menor si $\theta = \pi$

Si $\theta = \frac{\pi}{2}$ no tienen correlación

etc

$B^t B$ matriz de covarianzas



Queremos

$$\min \sum_{i=1}^{n} \|r_i\|^2$$

Dados $\bar{a}_1 \dots \bar{a}_m \in \mathbb{R}^n$

$$\underbrace{\|\bar{a}_i\|^2}_{\text{datos}} = \|r_i\|^2 + \|z_i\|^2$$

$$\min \sum \|r_i\|^2$$

equivale a

$$\underbrace{\max \sum \|z_i\|^2}_{\text{máx varianza en la proyección.}}$$

Es decir, la recta conserva la variabilidad!

Y esta es una propiedad esperable!