Aplicaciones de la SVD (ejemplos)

(1) Compresión de datos

Una matriz de rango r se puede escribir como suma de r matrices de rango 1:

$$A = U\Sigma V^* = \sum_{k=1}^{r} \sigma_k U_k V_k^*$$

U_k = k-ésima columna de U

V_k^* = k-ésima fila de V^*

Se puede hacer la cuenta pero es más facil verificarlo directamente en la base $\{v_1, .., v_n\}$ de K^n:

Sabemos que $A v_j = U\Sigma V^* v_j = U \Sigma e_j = U \sigma_j e_j = \sigma_j e_j$, $1 \le j \le r$

y $A v_j = 0$, $r+1 \le j \le n$

Por otro lado

$$\sum_{k=1}^{r} \sigma_k U_k V_k^* \cdot v_j = \begin{cases} \sigma_j U_j & \text{para } 1 \le j \le r \quad (\text{pues } V_k^* \cdot v_j = \delta_{kj}) \\ 0 & \text{para } r+1 \le j \le n. \end{cases}$$

Coinciden!

¿Por qué compresión de datos?

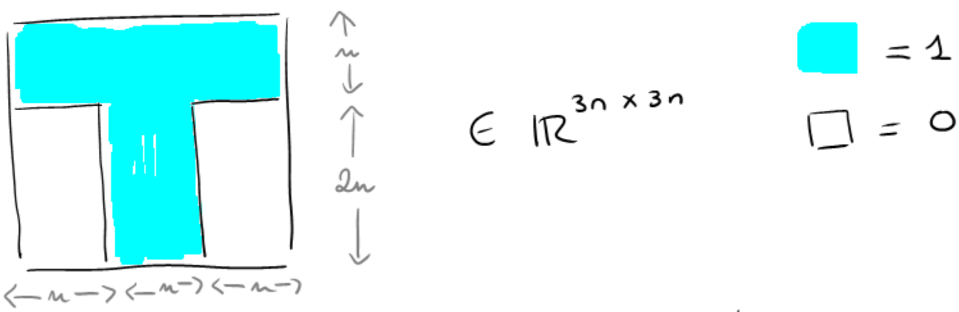
En principio la matriz A tiene $m \times n$ coeficientes

Pero para guardar cada $U_k V_k^*$ se necesitan sólo $m+n$ coeficientes

y luego se suman r así $\Rightarrow$ $r(m+n) + r$ datos ...

Y $r(m+n) << mn$ si $rg(A)$ chico y m, n grandes por ejplo!

Ejemplo

Sea $A =$  $\in \mathbb{R}^{3n \times 3n}$

A tiene claramente rango 2, luego se puede escribir como suma de 2 matrices de rango 1: Se comprimen (en forma sistemática) los $9n^2$ coeficientes de A en

- 2 valores singulares no nulos
- 2 vectores u_1, u_2: $6n$ coeficientes
- 2 vectores v_1, v_2. $6n$ coeficientes.

O sea $12n + 2$ contra $9n^2$ coeficientes

Por ejplo para $n = 100$: 1202 contra 90000 ......

Aquí: $A^t A =$  $= \begin{pmatrix} n & n & n \\ n & 3n & n \\ n & n & n \end{pmatrix}$

cumple $N(A^tA) = N(A) = \langle \underbrace{e_2-e_1, \dots, e_n-e_1}_{n-1}, \underbrace{e_{2n+1}-e_1, \dots, e_{3n}-e_1}_{n}, \underbrace{e_{2n+1}-e_{2n}, \dots, e_{3n-1}-e_{2n}}_{n-1} \rangle$

y $N(A^tA)^\perp = \langle \underbrace{(\underbrace{1,\dots,1}_{n}, \underbrace{0,\dots,0}_{n}, \underbrace{1,\dots,1}_{n})}_{w_1}, \underbrace{(\underbrace{0,\dots,0}_{n}, \underbrace{1,\dots,1}_{n}, \underbrace{0,\dots,0}_{n})}_{w_2} \rangle$ es invariante

Se tiene $A^tA\, w_1 = (2n^2, \dots, 2n^2) = 2n^2 w_1 + 2n^2 w_2$

$A^tA\, w_2 = (n^2, 3n^2, n^2) = n^2 w_1 + 3n^2 w_2$

luego ese bloque (restringido a $\langle w_1, w_2 \rangle$) es $\begin{bmatrix} 2n^2 & n^2 \\ 2n^2 & 3n^2 \end{bmatrix}$

con característico $\lambda^2 - 5n^2\lambda + 4n^4 = (\lambda - 4n^2)(\lambda - n^2)$

y autovectores $w_1 + 2w_2 = (1,\dots,1,2,\dots,2,1,\dots,1)$ asociado a $4n^2$

y $w_1 - w_2 = (1,\dots,1,-1,\dots,-1,1,\dots,1)$ asociado a n^2

Así $A = U\Sigma V^t$ con $\Sigma = \begin{bmatrix} 2n & & & \\ & n & & \\ & & 0 & \\ & & & 0 \end{bmatrix}$, $V = \begin{bmatrix} v_1 & v_2 & \cdots \end{bmatrix}$, $U = \begin{bmatrix} u_1 & u_2 & \cdots \end{bmatrix}$

$= \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t$

$= 2n\, u_1 v_1^t + n\, u_2 v_2^t$

donde $v_1 = \left(\underbrace{\frac{\sqrt{6n}}{6n}, \dots, \frac{\sqrt{6n}}{6n}}_{n}, \underbrace{\frac{\sqrt{6n}}{3n}, \dots, \frac{\sqrt{6n}}{3n}}_{n}, \underbrace{\frac{\sqrt{6n}}{6n}, \dots, \frac{\sqrt{6n}}{6n}}_{n}\right)$

$v_2 = \left(\underbrace{\frac{\sqrt{3n}}{3n}, \dots, \frac{\sqrt{3n}}{3n}}_{n}, \underbrace{-\frac{\sqrt{3n}}{3n}, \dots, -\frac{\sqrt{3n}}{3n}}_{n}, \underbrace{\frac{\sqrt{3n}}{3n}, \dots, \frac{\sqrt{3n}}{3n}}_{n}\right)$

y $u_1 = \dfrac{Av_1}{\sigma_1} = \left(\underbrace{\frac{\sqrt{6n}}{3n}, \dots, \frac{\sqrt{6n}}{3n}}_{n}, \underbrace{\frac{\sqrt{6n}}{6n}, \dots, \frac{\sqrt{6n}}{6n}}_{2n}\right)$

$u_2 = \dfrac{Av_2}{\sigma_2} = \left(\underbrace{\frac{\sqrt{3n}}{3n}, \dots, \frac{\sqrt{3n}}{3n}}_{n}, \underbrace{-\frac{\sqrt{3n}}{3n}, \dots, -\frac{\sqrt{3n}}{3n}}_{2n}\right)$

2 Aproximación a matrices de rango k

Teorema de Eckard-Young, 1936:

$A = U\Sigma V^*$ con $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix} \in \mathbb{C}^{m\times n}$, con $\sigma_1 \geq \dots \geq \sigma_r > 0$

Sea $\|A\|_F = \sqrt{tr(AA^*)}$ la norma Frobenius de A, y sea k fijo, $1 \leq k \leq r$:

Entonces la matriz A_k más cercana a A de rango k con esa distancia

es la matriz $A_k = U\Sigma_k V^*$ donde $\Sigma_k = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & \end{bmatrix} \in \mathbb{C}^{m\times n}$.

(Se truncan los valores singulares luego del k-ésimo)

Además observemos que dado que U, V^* son unitarias,

$$\|A\|_F^2 = tr(U \Sigma V^* V \Sigma^* U^*) = tr(U \Sigma \Sigma^* U^*) = tr(\Sigma \Sigma^*) = \sigma_1^2 + \dots + \sigma_r^2$$

y de la misma forma

$$\|A - A_k\|_F^2 = \sigma_{k+1}^2 + \dots + \sigma_r^2$$

Es decir $d_F(A, \mathcal{M}_k) = \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}$

matrices de rango k

Esto sirve por ejemplo para "eliminar ruido," errores de medición etc, estableciendo una línea de corte entre los valores singulares, los más chicos o por debajo de un umbral se consideran despreciables y se reemplazan por 0. Luego la matriz "medida" A corresponde en realidad a la matriz real A_k y se puede establecer el "rango efectivo de A"

- Con otras normas vale lo mismo. En particular para la norma operador $\|A\|_2 = \sup_{x \in \mathbb{C}^n} \frac{\|Ax\|_2}{\|x\|_2} = \sup_{\|x\|_2 = 1} \|Ax\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2$

(donde $\|x\|_2$, $\|Ax\|_2$ es la norma usual de $\mathbb{C}^n$ y $\mathbb{C}^m$ respectivamente)

Vale $d_2(A, \mathcal{M}_k) = \|A - A_k\|_2 = \sigma_{k+1}$ (Ver ejercicio 13 de la práctica)

con esta norma

3 "Condición" de una matriz

La condición de A para el problema de resolución de un sistema lineal cuadrado $Ax = b$ mide cuán mal se comporta la solución x' cuando se midió mal b y se resolvió $Ax' = b'$ en lugar de $Ax = b$.

Para A inversible se evalúa cuán grande puede ser el error relativo $\frac{\|x' - x\|}{\|x\|}$ en términos del error relativo $\frac{\|b' - b\|}{\|b\|}$

Fijada una norma como $\|A\|_F$ ó $\|A\|_2$, se tiene:

$$Cond_F(A) := \|A\|_F \|A^{-1}\|_F \quad (\text{resp } Cond_2(A) := \|A\|_2 \|A^{-1}\|_2)$$

para A inversible, y $Cond(A) = \infty$ para A no inversible

satisface en cada caso $\frac{\|x' - x\|}{\|x\|} \leq Cond(A) \frac{\|b' - b\|}{\|b\|}$ y se alcanza

Se tiene $\mathrm{Cond}_2(A) = \frac{\sigma_1}{\sigma_n}$ si $A \in GL(n, \mathbb{C})$ Ver Ej. 13 Práctica

Las matrices mal puestas son las que son no inversibles, ya que perturbándolas un poco cambia drásticamente el comportamiento c/r a las soluciones

En 1936, Eckard y Young demostraron (para $\| \ \|_F$ en realidad)

$$d_F(A, \mathcal{M}_0) = \frac{\|A\|_F}{\mathrm{Cond}_F(A)} \quad \left(\text{donde } \mathcal{M}_0 = \{ A \in \mathbb{C}^{n \times n} : \det(A) = 0 \} \right)$$

matrices no inversibles

4 Inversa de Moore - Penrose y SVD (Ver Ej. 16 de la práctica)

Sea $A \in \mathbb{C}^{m \times n}$, no necesariamente cuadrada

Existen definiciones de matrices "pseudo-inversas" que son matrices lo más cercanas a lo que podría ser una inversa.

La más famosa es la pseudo-inversa de Moore - Penrose $A^{\dagger}$ de A

(descrita por E.H. Moore, 1920, Arne Bjerhammar, 1951, y R. Penrose, 1955)

que es una matriz $A^{\dagger} \in \mathbb{C}^{n \times m}$ que satisface

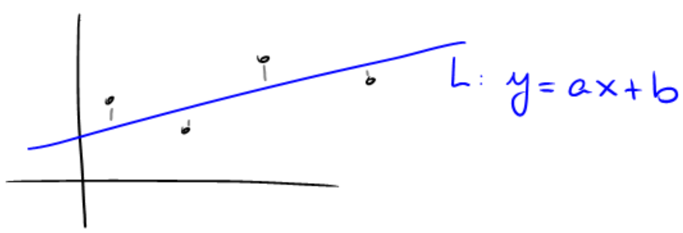
- $AA^{\dagger}A = A$ y $A^{\dagger}AA^{\dagger} = A^{\dagger}$
- $(AA^{\dagger})^* = AA^{\dagger}$ y $(A^{\dagger}A)^* = A^{\dagger}A$

Se puede probar que esa matriz $A^{\dagger}$ existe y es única, y se calcula fácil usando SVD:

Si $A = U \Sigma V^*$, entonces $A^{\dagger} = V \Sigma^{\dagger} U^*$ donde para

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix} \in \mathbb{C}^{m \times n}, \text{ se tiene } \Sigma^{\dagger} = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \end{bmatrix} \in \mathbb{C}^{n \times m}.$$

5 Y cuadrados mínimos (Ver Ejercicio 16 de la práctica)



¿Cuál es la recta $L: y = ax + b$ que se "ajusta mejor" a los ptos $(x_1, y_1), \dots, (x_n, y_n)$ del plano dados?

Donde "ajusta mejor" se refiere a minimizar la suma de los cuadrados de las distancias de y_i a $(ax_i + b)$: $\min \left(\sum_{i=1}^{n} \| y_i - (ax_i + b) \|^2 \right)$

El problema se puede plantear como resolver

$$\underbrace{\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}}_{A} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$ que en general no tiene solución.

Se puede probar que la solución es dada por $\begin{pmatrix} a \\ b \end{pmatrix} = A^{+} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$,

y esto se generaliza a pbs de cuadrados mínimos de dimensión mayor.

Etc, etc, etc... Hay muchas aplicaciones más...

Endomorfismos ortogonales y unitarios

Hemos visto que $U \in \mathbb{C}^{n\times n}$ es unitaria si sus columnas son una base ortonormal de $\mathbb{C}^n$, o equivalentemente si U es inversible y $U^{-1} = U^*$.

Si miramos $f_U : \mathbb{C}^n \to \mathbb{C}^n$, esto significa que f_U manda la base canónica de $\mathbb{C}^n$ a una base ortonormal de $\mathbb{C}^n$ (dada por las columnas de U), o también, dado que en la base canónica $\mathcal{E}$, $[f_U^*]_{\mathcal{E}} = [f_U]_{\mathcal{E}}^* = U^*$, que f_U es inversible y $f_U^{-1} = f_U^*$ (pues $[f_U^*]_{\mathcal{E}} = U^{-1} = [f_U]_{\mathcal{E}}^{-1}$)

Más aún, $f_U \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n)$ preserva el producto escalar, y la norma:

pues $\langle f_U(v), f_U(w) \rangle = \langle v, f_U^* f_U(w) \rangle = \langle v, w \rangle$, $\forall v, w \in \mathbb{C}^n$.

O sea en el caso real, donde el ángulo proviene del p.i, f_U preserva ángulos y distancias. Es un ejemplo de lo que se llama una isometría.

De hecho, todas estas propiedades caracterizan lo que se llama los endomorfismos unitarios (en el caso $K = \mathbb{C}$) u ortogonales (para $K = \mathbb{R}$) al menos en dimensión finita.

Definición (Endomorfismo unitario)

Sea $(V, \langle , \rangle)$ un K-e v con p.i, y sea $f \in \mathrm{End}_K(V)$.

Se dice que f es unitaria si f preserva el p.i, es decir

$$\langle f(v), f(w) \rangle = \langle v, w \rangle, \quad \forall v, w \in V$$

(En el caso euclídeo, i.e. $K = \mathbb{R}$, se dice que f es ortogonal)

Observación: p.i. vs. norma

En un $\mathbb{K}$-e.v V con p.i, $f \in \mathrm{End}_{\mathbb{K}}(V)$ preserva $\langle\,,\,\rangle$ sii f preserva $\|\ \|$. I.e.

$$\langle f(v), f(w)\rangle = \langle v, w\rangle,\ \forall v, w \in V \iff \|f(v)\| = \|v\|,\ \forall v \in V$$

Demostración

$(\Rightarrow)$ Es obvio. Para $(\Leftarrow)$ recuperemos $\langle\,,\,\rangle$ de $\|\ \|$:

$\forall v, w \in V$, $\langle v, w\rangle = \mathrm{Re}\langle v, w\rangle + i\,\mathrm{Im}\langle v, w\rangle$

y $\langle f(v), f(w)\rangle = \mathrm{Re}\langle f(v), f(w)\rangle + i\,\mathrm{Im}\langle f(v), f(w)\rangle$

Se tiene:

$\|v+w\|^2 = \langle v+w, v+w\rangle = \|v\|^2 + \langle v, w\rangle + \langle w, v\rangle + \|w\|^2 = \|v\|^2 + \|w\|^2 + 2\mathrm{Re}\langle v, w\rangle$

y del mismo modo $\|f(v+w)\|^2 = \|f(v)\|^2 + \|f(w)\|^2 + 2\mathrm{Re}\langle f(v), f(w)\rangle$

Esto implica, dado que f preserva la norma, que $\mathrm{Re}\langle f(v), f(w)\rangle = \mathrm{Re}\langle v, w\rangle$

Análogamente, para recuperar $\mathrm{Im}\langle f(v), f(w)\rangle = \mathrm{Im}\langle v, w\rangle$ en el caso complejo, y observando que $\mathrm{Im}(z) = -\mathrm{Re}(iz)$ para $z \in \mathbb{C}$:

$\|v + iw\|^2 = \langle v+iw, v+iw\rangle = \|v\|^2 + \bar{i}\langle v, w\rangle + i\langle w, v\rangle + \|w\|^2$

$= \|v\|^2 + \|w\|^2 - i\langle v, w\rangle - \overline{i\langle v, w\rangle} = \|v\|^2 + \|w\|^2 - 2\mathrm{Re}(i\langle v, w\rangle)$

y $\|f(v+iw)\|^2 = \|f(v)\|^2 + \|f(w)\|^2 - 2\mathrm{Re}(i\langle f(v), f(w)\rangle)$

Así $-\mathrm{Re}(i\langle v, w\rangle) = -\mathrm{Re}(i\langle f(v), f(w)\rangle) \Rightarrow \mathrm{Im}\langle f(v), f(w)\rangle = \mathrm{Im}\langle v, w\rangle$ $\boxtimes$

Proposición (transformaciones unitarias e inversa)

Sea $(V, \langle\,,\,\rangle)$ un $\mathbb{K}$-e.v con p.i. de dim. finita y sea $f \in \mathrm{End}_{\mathbb{K}}(V)$.

Entonces f es unitaria (ortogonal) $\iff$ f es inversible y $f^{-1} = f^*$

(Luego también $\iff$ $[f]_{\mathcal{B}}$ es unitaria para $\mathcal{B}$ base ortonormal de V)

Demostración

$(\Leftarrow)$ qpq $\langle f(v), f(w)\rangle = \langle v, w\rangle$, $\forall v, w \in V$.

Pero $\langle f(v), f(w)\rangle = \langle v, f^* f(w)\rangle \underset{f^* = f^{-1}}{=} \langle v, \mathrm{id}_V(w)\rangle = \langle v, w\rangle$

$(\Rightarrow)$ qpq f es inversible y $f^{-1} = f^*$, o sea qpq $f^* \circ f(w) = w$, $\forall w \in V$

(por estar en dim. finita, $f^* \circ f = \mathrm{id}_V \Rightarrow f \circ f^* = \mathrm{id}_V$ también)

Pero $\langle v, f^* \circ f(w)\rangle = \langle f(v), f(w)\rangle$ por def de f^*

$= \langle v, w\rangle$ por ser unitaria, $\forall v \in V$, $\forall w \in V$

Luego $f^* \circ f(w) = w$, $\forall w \in V$, considerando por ejemplo $v = f^* \circ f(w) - w$:

$\|f^* \circ f(w) - w\|^2 = 0 \Rightarrow f^* \circ f(w) - w = 0 \Rightarrow f^* \circ f(w) = w$, $\forall w \in V$.

La segunda equivalencia es simplemente porque sobre una base o.n. $\mathcal{B}$

$[f^*]_{\mathcal{B}} = [f]_{\mathcal{B}}^*$: $\quad f^{-1} = f^* \iff [f]_{\mathcal{B}}^{-1} = [f]_{\mathcal{B}}^*$ $\boxtimes$

Ejercicio Probar que f unitaria $\iff$ f manda base o.n en base o.n (en dim. finita)