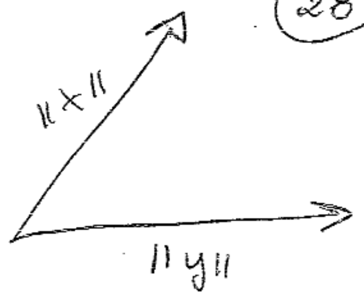
Gli spazi reali $\mathbb{R}^n$

$x, y \in \mathbb{R}^n$ $\quad \cos\theta = \dfrac{\langle x, y \rangle}{\|x\| \, \|y\|} = \dfrac{x^T y}{\|x\| \, \|y\|}$



Linear Correlation

(similarità del coseno)

Supponiamo di aver condotto un esperimento e che le osservazioni siano state salvate in due vettori

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Indichiamo con $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$

Problema Determinare l'esistenza di una relazione lineare di x e y ovvero determinare con quale grado le componenti di y sono correlate (linearmente) con quelle di x

Cioè vogliamo determinare (misurare) quanto y è vicino a una combinazione lineare di x ovvero se possiamo scrivere

$$y = \beta_0 \mathbb{1} + \beta_1 x \quad \text{per qualche } \beta_0 \text{ e } \beta_1$$

Soluzione Il coseno dell'angolo formato da x e y ci fornisce il grado di correlazione lineare tra questi vettori. Per dimostrare questa affermazione consideriamo

$$\mu_x = \frac{\sum_i x_i}{n} = \frac{\mathbb{1}^T x}{n}$$ media (tendenza principale dei dati)

$$\sigma_x = \sqrt{\frac{\sum (x_i - \mu_x)^2}{n}} = \frac{\|x - \mu_x \mathbb{1}\|_2}{\sqrt{n}}$$ deviazione standard: misura il grado con cui i dati sono sparpagliati

È frequente nelle applicazioni trovarsi di fronte a dati che siano stati calcolati con diverse unità di misura. Pertanto tali dati potrebbero essere difficilmente comparabili. Si procede quindi ad effettuare la cosiddetta "standardizzazione" in quantità indipendenti dall'unità di misura

DEF La standardizzazione di un vettore $x \in \mathbb{R}^n$ $\sigma_x \neq 0$ è data da

$$z_x = \frac{x - \mu_x \mathbb{1}}{\sigma_x}$$

Le componenti del vettore z_x si dicono z-scores (o punteggi standard).

I vettori standardizzati sono tali che $\|z\| = \sqrt{n}$ $\mu_z = 0$ $\sigma_z = 1$

Si verifica che:

$x, y \in \mathbb{R}^n$ t.c. $\sigma_x \neq 0$ e $\sigma_y \neq 0$

$z_x = z_y \iff \exists \beta_0, \beta_1$ t.c. $y = \beta_0 \mathbb{1} + \beta_1 x$ $\beta_1 > 0$

$z_x = -z_y \iff \exists \beta_0, \beta_1$ t.c. $y = \beta_0 \mathbb{1} + \beta_1 x$ $\beta_1 < 0$

In altre parole $y = \beta_0 \mathbb{1} + \beta_1 x$ per qualche β_0 e β_1 $\iff$ $z_x = \pm z_y$ (e diremo che y è perfettamente correlato con x)

Poiché z_x dipende da x, z_x sarà tanto più vicino a $\pm z_y$ quanto più x è linearmente correlato a y.

Osserviamo che $\|z_x\| = \sqrt{n} = \|\pm z_y\| \Rightarrow z_x$ e $\pm z_y$ differiscono solo nell'orientazione quindi una misura naturale per valutare quanto z_x sia vicino a $\pm z_y$ è data dal $\cos\theta$

$$\rho_{xy} = \cos\theta = \frac{z_x^T z_y}{\|z_x\| \, \|z_y\|} = \frac{z_x^T z_y}{n} = \frac{1}{n} \frac{(x - \mu_x \mathbb{1})^T (y - \mu_y \mathbb{1})}{\sigma_x \qquad \sigma_y}$$

$$= \frac{(x - \mu_x \mathbb{1})^T (y - \mu_y \mathbb{1})}{\|x - \mu_x \mathbb{1}\|_2 \, \|y - \mu_y \mathbb{1}\|}$$

29 bis

ρ_{xy} è detto coefficiente di correlazione lineare

$\rho_{xy} = 0 \quad x \perp y \Rightarrow x, y$ sono completamente scorrelati

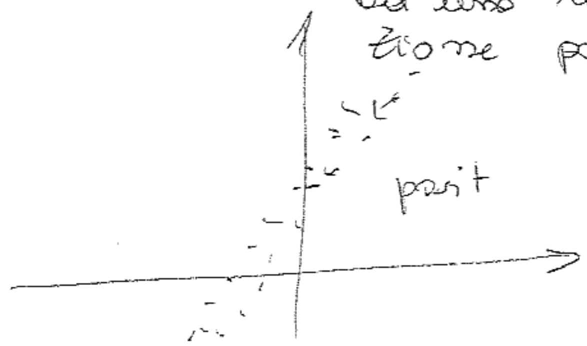
$|\rho_{xy}| = 1 \iff y$ è perfettamente correlato a x

y positivam. correlato a x se $\beta_1 > 0$

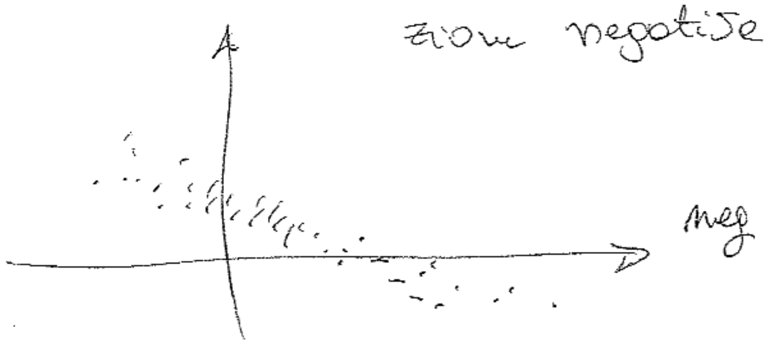
y negativam. " " " " $\beta_1 < 0$

$|\rho_{xy}|$ misura il grado con cui y è lin. legato a x

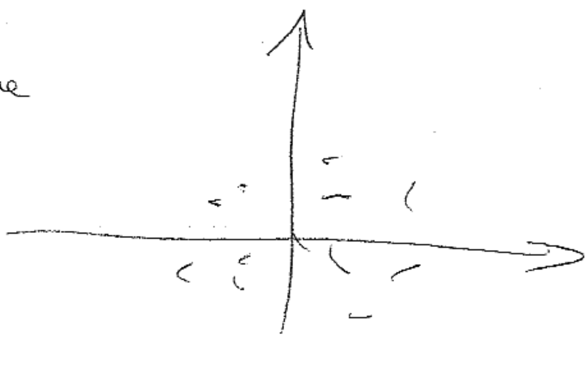
$\rho_{xy} \approx 1$ correlaz. positiva ovvero i dati appartengono ad una retta con inclinazione positiva



$\rho_{xy} \approx -1$ correlaz. negativa $\Rightarrow$ i dati giacciono su una retta con inclinazione negativa



$\rho_{xy} \approx 0$ nessuna correlazione



Decomposizione a valori singolari

Sia $A \in \mathbb{C}^{m\times n}$ $\text{rank}(A) = p$ ($r = \min(m,n)$), allora $\exists\, U, \Sigma, V$

$U \in \mathbb{C}^{m\times m}$ unitaria $(U^H U = I_m = U U^H)$

$V \in \mathbb{C}^{n\times n}$ unitaria $(V^H V = I_n = V V^H)$

e Σ matrice diagonale a blocchi $\Sigma \in \mathbb{R}^{m\times n}$ $\Sigma = \begin{pmatrix} \overset{p\times p}{D} & 0 \\ 0 & 0 \end{pmatrix}$

$D \in \mathbb{R}^{k\times k}$ $D = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$ con $\overbrace{\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_p > 0}^{\text{ordinati, positivi}}$

$p \leq \min(m, n)$ NO autovalori

σ_i sono detti valori singolari di A

$A = U \Sigma V^H$ dec. a valori singolari (SVD) di A

$U = [u_1 | u_2 | \dots | u_m]$ u_i si dicono vettori singolari sinistri

$V = [v_1 | v_2 | \dots | v_n]$ v_i " " " " destri

Si osservi che dalla relazione

$$A = U\Sigma V^H = \boxed{\sum_{i=1}^{r} \sigma_i u_i v_i^H}$$

$v_1, \dots, v_r$: basi ortonormali per spazio righe
$u_1, \dots, u_r$: base ortonormale per spazio colonne
$v_{r+1}, \dots, v_n$: base ortonormale per spazio nullo
$u_{r+1}, \dots, u_m$: base ortonormale per $N(A^T)$
$r = \text{rank}(A)$

inoltre da

$$AV = U\Sigma \quad \Rightarrow \quad A v_i = \sigma_i u_i \qquad i = 1, \dots, r$$

$$A^H = V \Sigma^H U^H \quad \Rightarrow \quad A^H U = V \Sigma^H \quad \Rightarrow \quad A^H u_i = \sigma_i v_i \qquad i = 1, \dots, r$$

<u>Teorema</u> Sia $A \in \mathbb{C}^{m\times n}$ e sia $A = U\Sigma V^H$ la SVD di A

t.c. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_p = 0$ $(\text{rank}(A) = k)$

Allora si ha che

1) $$A = U_k \Sigma_k V_k^H = \sum_{i=1}^{k} \sigma_i u_i v_i^H$$ dove

$U_k \in \mathbb{C}^{m\times k}$ è la matrice $U = [u_1 | u_2 | \dots | u_k]$ dei primi k vettori singolari sinistri

$V_k \in \mathbb{C}^{k\times n}$ è la matrice $V = [v_1 | v_2 | \dots v_k]$ le cui colonne sono i primi k vett. sing destri

$\Sigma_k \in \mathbb{R}^{k \times k}$ è la matrice diagonale i cui elementi diagonali sono $\sigma_1, \sigma_2, \dots, \sigma_k$

Infatti supponendo $n \le m$ (altrimenti si sostituisca A con A^H) possiamo partizionare Σ a blocchi nel seguente modo

$$\Sigma = \begin{array}{cc} \left[\begin{array}{cc} \Sigma_k & 0 \\ 0 & 0 \end{array}\right] & \begin{array}{l} k \text{ righe} \\ m-k \text{ righe} \end{array} \\ \begin{array}{cc} k \text{ colonne} & n-k \text{ colonne} \end{array} & \end{array}$$

e partizionare le matrici U e V^* nel seguente modo

$$U = [\overbrace{u_1 | \dots | u_k}\, | u_{k+1} | \dots | u_m] = [U_k \,|\, \tilde{U}^*_{m-k}]$$

$$V = [v_1 | \dots | v_k | \underbrace{v_{k+1} | \dots | v_n}] = [\underset{n \times k}{V_k} \,|\, \underset{n \times n-k}{\tilde{V}^*_{n-k}}]$$

dalla def di SVD risulta

$$A = U \Sigma V^H = U_k \Sigma_k V_k^H$$

2) $\mathrm{Null}(A) = \{x \in \mathbb{R}^n \,|\, Ax = 0\} \equiv \mathrm{span}\{v_{k+1}, v_{k+2}, \dots v_n\}$

Sia $x \in \mathbb{R}^n$ t.c. $Ax = 0 \iff U \Sigma V^H x = 0 \iff \Sigma V^H x = 0$

indichiamo con $\underset{m \times n}{z} = \underset{}{\Sigma}\, \underset{n \times n}{V^H} \underset{n \times 1}{x}$ questo vettore si può partizionare nel seguente modo $z \in \mathbb{C}^{m \times 1}$

$$z = \begin{array}{c} \\ k \\ m-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \left[\begin{array}{cc} \Sigma_k & 0 \\ 0 & 0 \end{array}\right] \end{array} \left[\begin{array}{c} V_k^H \\ V_{n-k}^H \end{array}\right] \begin{array}{l} k \times n \\ (n-k) \times n \end{array} x \in \mathbb{C}^{m \times 1}$$

$$= \left[\begin{array}{c} \Sigma_k V_k^H x \\ 0 \end{array}\right] \begin{array}{l} k \text{ componenti} \\ m-k \text{ componenti} \end{array}$$

quindi $\Sigma V^H x = 0 \Rightarrow \Sigma_k V_k^H x = 0 \Rightarrow V_k^H x = 0$

$\Rightarrow$ quest'ultima relazione ci dice che x è ortogonale alle prime k colonne di V

$\Rightarrow$ poiché V è $\boxed{\text{unitaria}}$ possiede inversa uguale alla trasposta $\Rightarrow x$ deve essere generato (comb. lineare) dalle restanti colonne di $V \Rightarrow x \in \mathrm{span}\{v_{k+1} \dots$

rette colonne $\quad m\times n \quad m\times 1$

3) $R(A) = \{ y \in \mathbb{R}^m \mid Ax = y \} = \text{span}\{u_1, u_2, \dots u_k\}$

Sia $y = Ax = \underbrace{U\Sigma V^H}x = U_k \Sigma_k \overbrace{V_k^H x}^{z} = U_k z$

$\Rightarrow y = z_1 u_1 + \dots + z_k u_k \Rightarrow y \in \text{span}\{u_1, \dots u_k\}$

4) $\text{rank}(A) = k$ (poiché $\text{rank}(A) = \dim R(A)$)

5) σ_i^2 sono gli autovalori di $A^H A$ $(\sigma_i^2 \in \sigma(A^H A))$

consideriamo $A^H A = (U\Sigma V^H)^H (U\Sigma V^H) =$

V sono gli autovettori di A^TA
U sono gli autovettori di AA^T

$= V\Sigma^H \underline{U^H U} \Sigma V^H = V \Sigma_k^2 V^H$

$A^H A$ è simile a Σ_k^2 mediante la trasformazione unitaria $V \Rightarrow \sigma(A^H A) = \sigma(\Sigma_k^2) = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \dots\}$

ricordare che traccia e $\rho(A)$ di matrici simili coincidono

6) $\|A\|_F^2 = \text{trace}(A^H A) = \sum_{i=1} \sigma_i^2$

(ricordare i concetti di norme $\| \|_F$ $\| \|_2$)

$\|A\|_2^2 \overset{\text{def}}{=} \rho(A^H A) = \sigma_1^2 \Rightarrow \|A\|_2 = \sigma_1$

$\|A\|_F^2 = \sum_{ij} |a_{ij}|^2 = \sum_i \|A_{i*}\|_2^2 = \sum_j \|A_{*j}\|_2^2 = \text{trace}(A^H A)$

Teor. TSVD (Truncated SVD)

$A \in \mathbb{C}^{m\times n}$ $\quad \text{rank}(A) = k$ $\quad r \in \mathbb{N}$ $\quad \boxed{r < k}$

$A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^H$ è la matrice "più vicina ad A" in $\| \|_2$ cioè

$$\min_{\substack{\text{rank}(B) = r \\ B \in \mathbb{C}^{m\times n}}} \|A - B\|_2 = \|A - A_r\|_2 = \sigma_{r+1}$$

$(r+1)$-esimo valore singolare

$A \in \mathbb{R}^{m \times n}$ $\quad$ rank$(A) = r$ (1)

$\exists U \in \mathbb{R}^{m \times m}$ $\quad \exists V \in \mathbb{R}^{n \times n}$

$A = U \Sigma V^T$ $\quad$ ortogonali t.c.

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & & & 0 \end{pmatrix} \qquad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

i valori singolari σ_i^2 sono gli autovalori di A^TA
ed i vettori singolari sono particolari autovettori di A^TA
Un algoritmo per il calcolo della SVD è dato dal metodo QR applicato ad A^TA (senza calcolo diretto del prodotto A^TA)

INTERPRETAZIONE GEOMETRICA

I valori singolari $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ di una matrice forniscono informazioni sulla quantità di distorsione che si ha utilizzando la trasformazione lineare definita dalla matrice A
Possiamo vedere ciò analizzando come A distorce una sfera

Supponiamo $A \in \mathbb{R}^{n \times n}$ non singolare e sia la sfera in $\mathbb{R}^n$ (sfera unitaria) $\boxed{S_2 = \{ x \mid \|x\|_2^2 = 1 \}}$
vettori di norma = 1
La natura dell'immagine $A(S_2)$ si può ottenere considerando la SVD di A

A è sempre invertibile ?

$$A = U\Sigma V^T \Rightarrow A^{-1} = V \Sigma^{-1} U^T$$

$y \in A(S_2) \Rightarrow \exists x \in \mathbb{R}^n \quad \|x\| = 1 \quad$ t.c. $\quad Ax = y$

$$1 = \|x\|_2^2 = \|A^{-1}Ax\|_2^2 = \|V\Sigma^{-1}U^T \underbrace{\cancel{Ax}}_{y}\|_2^2 =$$

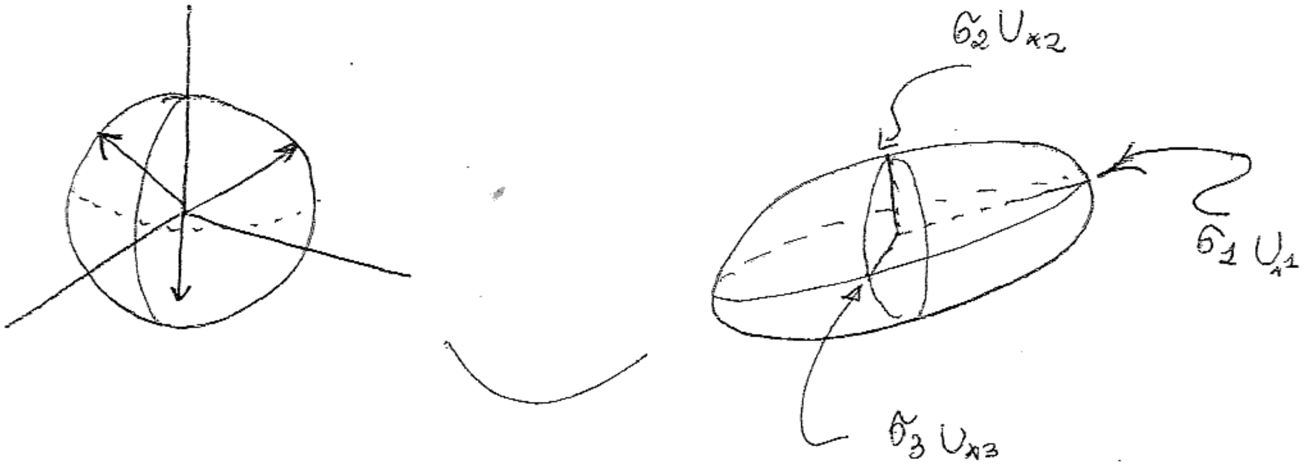
$$= \|\cancel{V}\, \Sigma^{-1} U^T y\|_2^2 = \|\Sigma^{-1} w\|_2^2 =$$
$w = U^T y$

$$= \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \dots + \frac{w_n^2}{\sigma_n^2}$$

$U^T A(S_2)$ è l'ellissoide il cui k-simo semiasse ha lunghezza σ_k. Poiché U^T è una matrice ortogonale quindi conserva la lungh.

U^T influisce solo sull'orientazione di $A(S_2)$ (2)

$\Rightarrow A(S_2)$ è un ellissoide il cui K-simo semiasse è σ_k



Il grado di distorsione di una sfera unitaria sotto la trasformazione lineare A è misurato da

$$\kappa_2 = \frac{\sigma_1}{\sigma_n}$$ il rapporto tra il più grande ed il più piccolo valore singolare

Inoltre il più lungo ed il più corto vettore dell'ellissoide $A(S_2)$ hanno rispettivamente lunghezza

$$\max_{\|x\|_2=1} \|Ax\|_2 = \|A\|_2 = \|U\Sigma V^T\|_2 = \|D\|_2 = \sigma_1$$

$$\min_{\|x\|_2=1} \|Ax\|_2 = \frac{1}{\|A^{-1}\|_2} = \frac{1}{\|V\Sigma^{-1}V^T\|_2} = \frac{1}{\|D^{-1}\|_2} = 1/\sigma_n$$

$$\boxed{\kappa_2 = \frac{\sigma_1}{\sigma_n}}$$ è chiamato numero di condizione in norma 2

Se $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ sono i valori singolari di A ($m \times n$) allora per ogni $K < r$ la distanza da A alla matrice di rango K più vicina ad A è data da (3)

$$\sigma_{K+1} = \min_{\text{rank}(B)=K} \| A - B \|_2$$

Filtraggio di dati

La SVD può essere utilizzata nelle applicazioni per effettuare un ordinamento di dati rumorosi ed estrarre informazioni rilevanti.

Supponiamo che A ($n \times m$) sia una matrice contenente dei dati contaminati da un certo livello di rumore.

La SVD permette di riscrivere A in r componenti mutuamente ortogonali

$$A = U \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots \sigma_r & \\ 0 & & 0 \end{pmatrix} V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sigma_i Z_i$$

$Z_i = u_i v_i^T$ e $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ (Z ha la stessa dimensione di A)

Le matrici $(Z_1, Z_2, \dots Z_r)$ costituiscono un insieme di matrici ortonormali perché (considerando la norma di Frobenius come prodotto scalare)
a vettori ortonormali

$$\langle Z_i, Z_j \rangle = \text{trace}(Z_i^T Z_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$\sigma_i = \text{trace}(Z_i^T A)$ può essere interpretato come la porzione di A che giace nella "direzione" di Z_i

In molte applicazioni il rumore che contamina i dati contenuti in A è random (non direzionale) nel senso che il rumore è distribuito più o meno uniformemente lungo ogni Z_i

Conseguentemente ogni termine $\sigma_i Z_i$ contiene approssimativamente lo stesso livello di rumore.

Questo significa che indicando con

$SNR(\sigma_i Z_i)$ il signal-to-noise-ratio in $\sigma_i Z_i$

si ha (si può definire il rapporto tra segnale-rumore come il reciproco del coefficiente di variazione ovvero il rapporto tra la media e la dev. standard del segnale o di una sua misurazione)

$$SNR(\sigma_1 Z_1) \geq SNR(\sigma_2 Z_2) \geq \dots \geq SNR(\sigma_r Z_r)$$

Se i valori singolari $\sigma_{k+1}, \sigma_{k+2}, \dots, \sigma_r$ sono piccoli relativamente a

$$\frac{(\text{rumore totale})}{r}$$

allora i termini $\sigma_{k+1} Z_{k+1}, \dots \sigma_r Z_r$ hanno un piccolo SNR.

Quindi eliminando alcuni termini $A = \sum_{i=1}^{r} \sigma_i Z_i$ sebbene perdiamo una piccola parte del segnale complessivo rimuoviamo una parte ampia delle componenti del rumore totale in A.

Questo spiega perché la TSVD $A_k = \sum_{i=1}^{k} \sigma_i Z_i$ può in molti casi filtrare il rumore in A senza una perdita <u>"significativa"</u> di informazioni

Per determinare il migliore valore di k di solito si applicano delle procedure empiriche che variano da applicazione ad applicazione.

Un meccanismo iniziale per scegliere k potrebbe anche essere basato sull'osservazione del gap tra il più grande ed il più piccolo valore singolare.

Come calcolare la ~~f~~ decomposizione SVD di A

Supponiamo $m \geq n$ (se fosse $m < n$ basta riferirsi alla matrice A^H)

1) Si calcolano gli autovalori e gli autovettori normalizzati in $\| \|_2$ di $A^H A$
e si considera la seguente decomp. in forma di Shur di $A^H A$

$$A^H A = Q D Q^H \qquad D \in \mathbb{R}^{n \times n} \qquad Q \in \mathbb{C}^{n \times n}$$

D matrice degli autovalori in ordine non crescente di $A^H A$ e Q è la matrice degli autovettori

2) Si calcola la matrice

$$C = AQ \in \mathbb{C}^{m \times n}$$

3) Si calcola la fattorizzazione QR di C utilizzando la tecnica del pivoting per ottenere una matrice triang. sup con elementi principali non negativi ed ordinati in modo non crescente

$$C\Pi = UR = U \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \qquad R_1 \in \mathbb{C}^{n \times n}$$

Π matrice di permutazione
$(R_1)_{ii} \geq 0$

$$A = CQ^H = C\Pi\Pi^H Q^H = UR\Pi^T Q^H$$

$$\Rightarrow A^H A = Q\Pi R^T U^H U R \Pi^T Q^H = Q\Pi R^T R \Pi^T Q^H$$

$$\Rightarrow \quad A^H A = Q \Pi R^H R \Pi^T Q^H = Q \Pi R_1^H R_1 \Pi^T Q^H$$

da $\quad A^H A = Q^H D Q \quad$ otteniamo

$$Q^H D Q^H = Q \Pi R_1^H R_1 \Pi^T Q^H$$

$$D = \Pi R_1^H R_1 \Pi^T$$

$$\Rightarrow \quad R_1^H R_1 = \underbrace{\Pi^T D \Pi}_{\text{diagonale}}$$

$\Rightarrow \quad R_1^H R_1$ è diagonale $\Rightarrow$ R_1 è diagonale

Si pone $\quad \Sigma = R \quad$ e $\quad V = Q \Pi$

e si ottiene la SVD di A.

Infatti

$$A = U R \Pi^T Q^H \equiv U \Sigma V^H$$

SVD Troncata (TSVD)

La SVD consente di risolvere il seguente problema di minimo

Data $A \in \mathbb{C}^{m \times n}$ ~~$\operatorname{rank}(A) = K$~~ ed $r \in \mathbb{N}$ $\quad r < K$

determinare la matrice $B \in \mathbb{C}^{m \times n}$ $\operatorname{rank}(B) = r$

t.c. $\quad \min \|A - B\|_2 \qquad$ più "vicina" ad A

Prop. $A \in \mathbb{C}^{m\times n}$ $\quad rank(A) = K$

$$A = U \Sigma V^H$$

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > \sigma_{k+1} = \sigma_{k+2} \dots = \sigma_p = 0$$

$$\min_{\substack{B \in \mathbb{C}^{m\times n} \\ rank(B) = r}} \| A - B \|_2 = \| A - A_r \|_2 = \sigma_{r+1}$$

con $\quad A_r = \sum_{j=1}^{r} \sigma_j u_j v_j^H$

Applicazione della SVD al problema dei minimi quadrati

Consideriamo il sist. lineare

$$Ax = b$$

$A \in \mathbb{C}^{m\times n}$ $\qquad m \geq n$ $\quad$ più righe che colonne

Se il sist. non ha soluzione fissata una norma vettoriale $\|\cdot\|$ si cercano i vettori $x \in \mathbb{C}^n$ s.t.

$$\min_x \| Ax - b \|_2$$

Questo problema si chiama "problema ai minimi quadrati"

Avete visto come la soluz. minima si possa calcolare utilizzando il sistema delle eq. normali

$$\boxed{A^H A x = A^H b}$$

vediamo come è possibile calcolare la soluzione del problema ai minimi quadrati utilizzando la fatt. SVD

Sia $A \in \mathbb{C}^{m\times n}$ con $rank(A) = K$. $m \geq n \geq K$ Consideriamo

$$A = U \Sigma V^H \qquad U \in \mathbb{C}^{m\times m} \quad V \in \mathbb{C}^{n\times n} \qquad \Sigma \in \mathbb{R}^{m\times n} \quad \Sigma = (\sigma_1 \dots \sigma_k\, 0 \dots 0)$$

$$\|Ax-b\|_2^2 \underset{\downarrow}{=} \|U^H(Ax-b)\|_2^2 = \|U^H A V V^H x - U^H b\|_2^2$$

$\|\|_2$ inv. x trasf. ortog.

$$\underline{\underline{y = V^H x}} \quad \|U^H A V y - U^H b\|_2^2$$

$$\overset{\text{dalla SVD}}{=} \|\Sigma y - U^H b\|_2^2 \qquad y = V^H x \in \mathbb{C}^n$$

$U = [u_1 | u_2 | \dots u_m]$

Osserviamo che

$$\Sigma y = \begin{bmatrix} \sigma_1 y_1 \\ \sigma_2 y_2 \\ \vdots \\ \sigma_k y_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} 1 \\ \\ \\ \\ \\ \\ m \end{matrix} \qquad U^H b = \begin{bmatrix} u_1^H \\ u_2^H \\ \vdots \\ u_m^H \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ b_{n+1} \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} u_1^H b \\ u_2^H b \\ \vdots \\ u_n^H b \\ \hline \vdots \\ u_m^H b \end{bmatrix}$$

$$\|U^H b\|_2^2 = \sum_{i=1}^{n} (u_i^H b)^2 + \sum_{i=n+1}^{m} (u_i^H b)^2$$

$$= \sum_{i=1}^{K} (u_i^H b)^2 + \sum_{i=K+1}^{m} (u_i^H b)^2$$

$$\Sigma y - U^H b = \begin{bmatrix} \sigma_i y_i - u_i^H b \\ \\ - u_i^H b \end{bmatrix} \begin{matrix} i=1 \\ \vdots \\ k \\ k+1 \\ \vdots \\ m \end{matrix}$$

$$\|\Sigma y - U^H b\|_2^2 = \sum_{i=1}^{K} |\sigma_i y_i - u_i^H b|^2 + \sum_{i=K+1}^{m} |u_i^H b|^2$$

Il minimo della precedente espressione viene raggiunto quando il primo addendo è nullo

$$\sigma_i y_i - u_i^H b = 0 \qquad y_i = \frac{u_i^H b}{\sigma_i} \quad i = 1, \dots K$$

$$y_i = 0 \qquad i = K+1, \dots n$$

Il vettore di minima norma

y^* è

$$y_i^* = \begin{cases} \dfrac{u_i^H b}{\sigma_i} & i = 1, \dots k \\ 0 & i = k+1, \dots, n \end{cases}$$

$y = V^H x \Rightarrow$

$x = Vy$ e $\|x\|_2 = \|y\|_2$ la sol. del prob. ai min 2

$$x^* = Vy^* = \sum_{i=1}^{k} y_i^* v_i = \sum_{i=1}^{k} \frac{u_i^H b}{\sigma_i} v_i$$

ed

$$\|Ax^* - b\|_2^2 = \sum_{i=k+1}^{m} |u_i^H b|^2$$

Decomposizione a valori singolari di $A \in \mathbb{C}^{m\times n}$

$U = [u_1 \dots u_m]$ $V = [v_1 \dots v_n]$

$A \in \mathbb{R}^{m\times n}$ $\exists U \in \mathbb{R}^{m\times m}$ $\exists V \in \mathbb{R}^{n\times n}$

U V ortogonali

$\exists \Sigma$ $m\times n$ diagonale sup a blocchi

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ 0 & & \ddots \end{bmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$

$r = \text{rango}(A)$

$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ $r \leq \min(m, n)$

$D = \text{diag}[\sigma_1 \dots \sigma_r)$

tale che

$$A = U \Sigma V^T$$

---

gli elementi di Σ si chiamano valori singolari di A

i vettori u_i vettori singolari sinistri

" v_i " " destri

/ U V non sono uniche mentre $\boxed{\Sigma \text{ è unica}}$

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^T$$

$$AA^T = U \Sigma V^T V \Sigma U^T = U \Sigma \Sigma^T U^T$$

U → autovettori AA^T

~~$\rho(AA^T) = \rho(\Sigma\Sigma^T)$~~ $A^T A = V (\Sigma^T \Sigma) V^T$

V → autovettori di $A^T A$

$$\|A\|_2 = \sqrt{\rho(A^T A)}$$

σ_i di A sono legati agli autovalori della matrice $A^T A$ $\longleftrightarrow$ $\sigma_i = \sqrt{\lambda_i (AA^T)}$

DEF. di Pseudoinverso

$A \in \mathbb{C}^{m \times n}$ $\qquad r = \text{rango}(A)$

$$\Sigma = \begin{array}{cc} r \times r & r \times (n-r) \\ \left[\begin{matrix} D & 0 \\ 0 & 0 \end{matrix}\right] & \\ (m-r) \times r & (m-r) \times (n-r) \end{array}$$

$A = U \Sigma V^T$

$$D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_r \end{bmatrix}$$

$$\Sigma \Sigma^+ \Sigma = \Sigma \quad \Rightarrow \quad \Sigma^T (\Sigma^+)^T \Sigma^+ = \Sigma^T$$

$$\Sigma^+ \Sigma \Sigma^+ = \Sigma^+$$

$$(\Sigma^+ \Sigma)^T = \Sigma^+ \Sigma$$

$$(\Sigma \Sigma^+)^T = \Sigma \Sigma^+$$

Si dice inverso generalizzato di A o inverso di Moore-Penrose

$$A^+ = V \Sigma^+ U^T$$

1) $A A^+ A = U \Sigma V^T V \Sigma^+ U^T U \Sigma V^T$
$= U \Sigma \Sigma^+ \Sigma V = U \Sigma^+ V = A$

2) $A^+ A A^+ = A^+$

3) $(A A^+)^T = A A^+$

4) $(A^+ A)^T = A^+ A$

5) $A^T A A^+ = A^T$

se rango$(A) = n < m$

$$A^+ = (A^T A)^{-1} A^T$$

rango$(A) = n = m$

$$A^+ = A^{-1}$$