

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

**LA SVD E LA DECOMPOSIZIONE POLARE.
ANALISI E APPLICAZIONI.**

Tesi di Laurea in Analisi Numerica

Relatore:
Chiar.ma Prof.ssa
VALERIA SIMONCINI

Presentata da:
CRISTINA MORRONI

II Sessione
Anno Accademico 2015/2016

*A tutti coloro che
mi hanno trascinato
fino a qui.*

Indice

Introduzione	7
Notazioni	9
1 SVD: la Decomposizione ai Valori Singolari	11
1.1 Esistenza e proprietà della SVD	11
1.2 Matrici a rango ridotto	16
1.3 Aspetti computazionali della SVD	18
2 La Decomposizione Polare	21
2.1 Esistenza della Decomposizione	21
2.2 Proprietà della Decomposizione	22
3 L'analisi di Procuste	25
3.1 Che cos'è l'analisi di Procuste	25
3.2 Un esempio di analisi di Procuste	28
A L'algoritmo MATLAB	33
Bibliografia	35

Introduzione

Scopo di questa tesi è presentare due decomposizioni matriciali molto usate in molteplici applicazioni: la Decomposizione ai Valori Singolari e la Decomposizione Polare.

In particolare, verrà approfondita la Decomposizione ai Valori Singolari: si tratta di una decomposizione molto generale, che si applica a una qualunque matrice anche rettangolare, ed è molto utile perchè, grazie ai valori singolari, dà a priori informazioni sulla matrice decomposta, quali il rango, il nucleo e il range della matrice. Viene usata inoltre per risolvere il problema ai minimi quadrati e per l'approssimazione della matrice con una di rango inferiore, in applicazioni quali l'elaborazione di immagini e di segnali.

La Decomposizione Polare, invece, generalizza la rappresentazione esponenziale dei numeri complessi ed è a sua volta usata in molti ambiti, dall'ingegneria aerospaziale all'analisi fattoriale. In particolare, verrà presentata l'applicazione di entrambe le decomposizioni all'analisi di Procuste, un metodo della statistica multivariata che, partendo da due configurazioni, le confronta numericamente determinando la trasformazione che, applicata ad una configurazione, minimizza la distanza tra le due.

La tesi si compone di tre capitoli: il primo riguarda la Decomposizione ai Valori Singolari, mostra la sua esistenza, varie proprietà e alcuni aspetti computazionali. Il secondo presenta la Decomposizione Polare, concentrandosi sulle proprietà del fattore ortogonale della decomposizione. Infine, nel terzo si introduce l'analisi di Procuste e viene mostrato come le decomposizioni

matriciali discusse si applicano ad essa. Si conclude con un esempio pratico di analisi di Procuste, illustrando anche l'algoritmo MATLAB utilizzato per ottenere i risultati.

Notazioni

Nelle pagine che seguono, faremo uso delle seguenti notazioni:

Definizione 0.1 (Norma-2 di un vettore). Sia $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$. Si definisce la *norma-2* o *norma euclidea* di \mathbf{x} come

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}.$$

Definizione 0.2 (Norma-2 di una matrice). Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$. Si definisce la *norma-2* o *norma euclidea indotta* di \mathbf{A} come

$$\|\mathbf{A}\|_2 = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \neq 0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \|\mathbf{x}\|_2=1}} \|\mathbf{Ax}\|_2.$$

Un'altra norma molto utile è la *norma di Frobenius*:

Definizione 0.3. Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$. Si definisce la *norma di Frobenius* di \mathbf{A} come

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{tr}(\mathbf{AA}^T)}$$

dove $\text{tr}(\mathbf{B})$ è la *traccia* della matrice $\mathbf{B} \in \mathbb{R}^{n \times n}$:

$$\text{tr}(\mathbf{B}) = \sum_{i=1}^n b_{ii}.$$

La norma di Frobenius ha, inoltre, una proprietà utile e immediata: se $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$,

$$\|\mathbf{A}\|_F^2 = \|\mathbf{a}_1\|_2^2 + \dots + \|\mathbf{a}_n\|_2^2 = \sum_{i=1}^n \|\mathbf{a}_i\|_2^2.$$

Si definisce, quindi, il *numero di condizionamento* di una matrice quadrata $\mathbf{A} \in \mathbb{R}^{n \times n}$ non singolare come

$$k(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|,$$

con $\|\cdot\|$ norma matriciale.

Useremo, inoltre, le seguenti nozioni:

Definizione 0.4. • $\mathbf{A} \in \mathbb{R}^{n \times n}$ è *ortogonale* se $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = I_n$, dove I_n è la matrice identità di ordine n .

• $\mathbf{A} \in \mathbb{R}^{n \times n}$ simmetrica è *definita positiva* se, per ogni $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$, si ha

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0.$$

• $\mathbf{A} \in \mathbb{R}^{n \times n}$ simmetrica è *semidefinita positiva* se, per ogni $\mathbf{x} \in \mathbb{R}^n$, si ha

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0.$$

Si dice, inoltre, che \mathbf{X} è la *radice quadrata* di una matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$ se

$$\mathbf{X}\mathbf{X} = \mathbf{A}.$$

Infine, useremo la notazione *matrici a rango ridotto* come traduzione per *rank-deficient matrices*.

Capitolo 1

SVD: la Decomposizione ai Valori Singolari

1.1 Esistenza e proprietà della SVD

La Singular Value Decomposition (SVD) [1], o Decomposizione ai Valori Singolari, è una fattorizzazione matriciale che permette di rappresentare una qualunque matrice come il prodotto di due matrici ortogonali e una diagonale.

Teorema 1.1.1 (Teorema di esistenza della SVD).

Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$. Allora esistono $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ ortogonali e $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ tali che

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad \text{con } \mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ \mathbf{0} & \dots & \mathbf{0} & \end{pmatrix}$$

dove $\sigma_1, \dots, \sigma_n \in \mathbb{R}$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$.

Le colonne $\{\mathbf{u}_i\}_{i=1, \dots, m}$ di \mathbf{U} sono dette *vettori singolari sinistri*, le colonne $\{\mathbf{v}_j\}_{j=1, \dots, n}$ di \mathbf{V} *vettori singolari destri* e i $\{\sigma_k\}_{k=1, \dots, n}$ *valori singolari*.

Dimostrazione. Osserviamo che non è restrittivo supporre $m \geq n$; infatti, se $n \geq m$, basta applicare il teorema ad \mathbf{A}^T .

Siano ora $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ vettori tali che $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$ e

$$\mathbf{Ax} = \sigma_1 \mathbf{y} \text{ con } \sigma_1 := \|\mathbf{A}\|_2.$$

Possiamo quindi trovare $\tilde{\mathbf{X}}_1 \in \mathbb{R}^{n \times (n-1)}$ e $\tilde{\mathbf{Y}}_1 \in \mathbb{R}^{m \times (m-1)}$ tali che

$$\mathbf{X}_1 := [\mathbf{x}, \tilde{\mathbf{X}}_1] \in \mathbb{R}^{n \times n} \text{ e } \mathbf{Y}_1 := [\mathbf{y}, \tilde{\mathbf{Y}}_1] \in \mathbb{R}^{m \times m}$$

siano ortogonali. Allora, poiché $\mathbf{y}^T \mathbf{Ax} = \mathbf{y}^T \sigma_1 \mathbf{y} = \sigma_1 \mathbf{y}^T \mathbf{y} = \sigma_1 \|\mathbf{y}\|_2^2 = \sigma_1$ e $\tilde{\mathbf{Y}}_1^T \mathbf{Ax} = \tilde{\mathbf{Y}}_1^T \sigma_1 \mathbf{y} = \sigma_1 \tilde{\mathbf{Y}}_1^T \mathbf{y} = 0$ (per come abbiamo definito \mathbf{Y}_1), avremo

$$\mathbf{Y}_1^T \mathbf{AX}_1 = \begin{pmatrix} \mathbf{y}^T \\ \tilde{\mathbf{Y}}_1^T \end{pmatrix} \mathbf{A} \begin{pmatrix} \mathbf{x} & \tilde{\mathbf{X}}_1 \end{pmatrix} = \begin{pmatrix} \sigma_1 & \mathbf{w}^T \\ \mathbf{0} & \mathbf{B} \end{pmatrix} =: \mathbf{A}_1$$

con $\mathbf{w} \in \mathbb{R}^{n-1}$ e $\mathbf{B} \in \mathbb{R}^{(m-1) \times (n-1)}$.

Ora:

$$\begin{aligned} \left\| \mathbf{A}_1 \begin{pmatrix} \sigma_1 \\ \mathbf{w} \end{pmatrix} \right\|_2^2 &= \left\| \begin{pmatrix} \sigma_1 & \mathbf{w}^T \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \mathbf{w} \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} \sigma_1^2 + \mathbf{w}^T \mathbf{w} \\ \mathbf{Bw} \end{pmatrix} \right\|_2^2 = \\ &= (\sigma_1^2 + \mathbf{w}^T \mathbf{w})^2 + \|\mathbf{Bw}\|_2^2 \geq (\sigma_1^2 + \mathbf{w}^T \mathbf{w})^2 \end{aligned}$$

quindi $\|\mathbf{A}_1\|_2^2 \geq \sigma_1^2 + \mathbf{w}^T \mathbf{w}$.

Ma $\sigma_1^2 = \|\mathbf{A}\|_2^2$ per definizione di σ_1 , e $\|\mathbf{A}\|_2 = \|\mathbf{A}_1\|_2$ (poiché \mathbf{A}_1 si ottiene da \mathbf{A} tramite trasformazioni ortogonali), quindi $\mathbf{w}^T \mathbf{w} = \|\mathbf{w}\|_2^2 = 0$;

allora $\mathbf{w} = \mathbf{0}$, quindi $\mathbf{A}_1 = \begin{pmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}$.

Iterando il procedimento su \mathbf{B} , otterremo $\mathbf{U} = \mathbf{Y}_1 \cdots \mathbf{Y}_n$, $\mathbf{V} = \mathbf{X}_1 \cdots \mathbf{X}_n$

$$\text{e } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}.$$

□

Osservazione 1. Per ogni $\mathbf{x} \in \mathbb{R}^n$ tale che $\|\mathbf{x}\|_2 = 1$, avremo

$$\|\mathbf{Ax}\|_2 = \|\mathbf{U}\Sigma\mathbf{V}^T\mathbf{x}\|_2 = \|\Sigma\mathbf{V}^T\mathbf{x}\|_2.$$

Se $\mathbf{y} := \mathbf{V}^T\mathbf{x}$, allora $\|\mathbf{y}\|_2 = 1$ (poiché \mathbf{V} è ortogonale) quindi

$$\|\Sigma\mathbf{V}^T\mathbf{x}\|_2 = \|\Sigma\mathbf{y}\|_2 = \left\| \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \dots & \mathbf{0} \end{pmatrix} \mathbf{y} \right\|_2 = \left\| \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \mathbf{y} \right\|_2 \leq \sigma_1,$$

che è il valore singolare massimo di \mathbf{A} ; esso viene raggiunto per $\mathbf{x} = \mathbf{v}_1$, infatti $\|\mathbf{Av}_1\|_2 = \|\mathbf{u}_1\sigma_1\|_2 = \sigma_1$ (poiché \mathbf{u}_1 , che è una colonna della matrice ortogonale \mathbf{U} , ha norma 1), quindi $\|\mathbf{A}\|_2 = \sigma_1$; questo fatto è anche evidente dalla dimostrazione.

Analogamente, si vede che per ogni $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{Ax}\|_2 \geq \sigma_n$ e $\|\mathbf{Av}_n\|_2 = \|\mathbf{u}_n\sigma_n\|_2 \geq \sigma_n$ quindi

$$\min_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = \sigma_n.$$

In base a tale osservazione, si definisce quindi il *numero di condizionamento* di una generica matrice $\mathbf{B} \in \mathbb{R}^{m \times n}$:

Definizione 1.1. Se $\mathbf{B} \in \mathbb{R}^{m \times n}$ tale che $\sigma_{\min\{m,n\}} \neq 0$, si definisce il *numero di condizionamento* di \mathbf{B} come

$$k_2(\mathbf{B}) := \frac{\sigma_1}{\sigma_{\min\{m,n\}}}.$$

Questo concetto generalizza quello di condizionamento per matrici quadrate non singolari: se $\mathbf{B} \in \mathbb{R}^{n \times n}$ è una matrice non singolare, il suo numero di condizionamento è definito come $k_2(\mathbf{B}) := \|\mathbf{B}\|_2 \|\mathbf{B}^{-1}\|_2$.

Scrivendo la \mathbf{B} ,

$$\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}^T \quad \text{con } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}$$

si ha $\|\mathbf{B}\|_2 = \sigma_1(\mathbf{B})$.

Ora $\mathbf{B}^{-1} = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$ con $\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_n \end{pmatrix}$ quindi

$$\|\mathbf{B}^{-1}\|_2 = \frac{1}{\sigma_n(\mathbf{B})} \quad \text{e} \quad k_2(\mathbf{B}) = \frac{\sigma_1(\mathbf{B})}{\sigma_n(\mathbf{B})}.$$

Osservazione 2. Se $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$ la sua norma di Frobenius sarà

$$\begin{aligned} \|\mathbf{A}\|_F^2 &= \text{tr}(\mathbf{A}^T \mathbf{A}) = \text{tr}(\mathbf{V}\mathbf{\Sigma}^T \mathbf{U}^T \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) = \text{tr}(\mathbf{\Sigma}^T \mathbf{\Sigma}\mathbf{V}^T \mathbf{V}) = \\ &= \text{tr} \left(\begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} \right) = \sum_{k=1}^n \sigma_k^2. \end{aligned}$$

Osservazione 3. La SVD è una generalizzazione della decomposizione in

autovalori; infatti, se $\mathbf{\Sigma}_1 := \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}$,

$$\mathbf{A}^T \mathbf{A} = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{V}\mathbf{\Sigma}^T \mathbf{U}^T \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T =$$

$$= \mathbf{V}\mathbf{\Sigma}^T \mathbf{\Sigma}\mathbf{V}^T = \mathbf{V}\mathbf{\Sigma}_1^2 \mathbf{V}^T \quad (\text{decomposizione spettrale di } \mathbf{A}^T \mathbf{A}).$$

Quindi $\sigma_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}^T \mathbf{A})}$, con λ_i autovalore di $\mathbf{A}^T \mathbf{A}$, $\lambda_i \geq 0$, per $i=1, \dots, n$.

In particolare, se $\mathbf{A} \in \mathbb{R}^{n \times n}$ simmetrica, $\sigma_i(\mathbf{A}) = |\lambda_i(\mathbf{A})|$.

Notiamo inoltre che, se $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$, $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n, \dots, \mathbf{u}_m]$ con $\mathbf{u}_i \in \mathbb{R}^m$ per $i = 1, \dots, m$ e $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ con $\mathbf{v}_j \in \mathbb{R}^n$

per $j = 1, \dots, n$, allora:

$$\mathbf{A} = (\mathbf{u}_1 \dots \mathbf{u}_n \dots \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ \mathbf{0} & \dots & \mathbf{0} & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} = \sum_{i=1}^n \mathbf{u}_i \sigma_i \mathbf{v}_i^T,$$

che è una somma di matrici di rango 1.

Quindi, se esiste $r \in \{1, \dots, n\}$ tale che $\sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0$, la decomposizione si limita ai termini fino all' r -esimo e il rango di \mathbf{A} risulta essere r .

Di seguito sono riportate alcune proprietà fondamentali della SVD:

Proposizione 1.1.2. *Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$, $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ la sua SVD. Siano $r \in \{1, \dots, n\}$ tale che $\sigma_r = \min\{\sigma_i | \sigma_i > 0\}$, $\mathbf{U}_1 = [\mathbf{u}_1, \dots, \mathbf{u}_r]$, $\mathbf{U}_2 = [\mathbf{u}_{r+1}, \dots, \mathbf{u}_m]$, $\mathbf{V}_1 = [\mathbf{v}_1, \dots, \mathbf{v}_r]$, $\mathbf{V}_2 = [\mathbf{v}_{r+1}, \dots, \mathbf{v}_n]$. Allora avremo che:*

1. \mathbf{A} ha rango massimo $\Leftrightarrow r=n$.
2. $\text{Range}(\mathbf{A}) = \text{Range}(\mathbf{U}_1)$.
3. $\text{Range}(\mathbf{A}^T) = \text{Range}(\mathbf{V}_1)$.
4. $\text{Ker}(\mathbf{A}) = \text{Range}(\mathbf{V}_2)$.
5. $\text{Range}(\mathbf{A}^T) = \text{Range}(\mathbf{U}_2)$.

Dimostrazione. 1. viene da quanto appena visto.

2. $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \Rightarrow \mathbf{A}\mathbf{V} = \mathbf{U}\Sigma$.

$$\mathbf{U}\Sigma = (\mathbf{U}_1 \ \mathbf{U}_2) \begin{pmatrix} \Sigma_1 \\ \mathbf{0} \end{pmatrix} = \mathbf{U}_1 \Sigma_1 \text{ con } \begin{pmatrix} \sigma_1 & & \mathbf{0} \\ & \ddots & \vdots \\ & & \sigma_r \ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{r \times n}, \text{ cioè lo}$$

spazio generato dalle colonne di \mathbf{A} coincide con quello generato dalle colonne di $\mathbf{U}_1 \Rightarrow \text{Range}(\mathbf{A}) = \text{Range}(\mathbf{U}_1)$.

3. Analogo al punto precedente, partendo da

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

4. Se $j \in \{r + 1, \dots, n\}$,

$$\mathbf{A}\mathbf{v}_j = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{v}_j = \mathbf{U}\mathbf{\Sigma} \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{pmatrix} \mathbf{v}_j = \mathbf{U}\mathbf{\Sigma} \begin{pmatrix} \mathbf{0} \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{pmatrix} = \mathbf{0}$$

quindi $\text{Ker}(\mathbf{A}) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\} = \text{Range}(\mathbf{V}_2)$.

5. Analogo al punto precedente, partendo da $\mathbf{A}^T\mathbf{U} = \mathbf{V}\mathbf{\Sigma}^T$.

□

1.2 Matrici a rango ridotto

Molti teoremi dell'Algebra lineare si basano sull'ipotesi di matrici di rango massimo; spesso, tuttavia, si ha a che fare con matrici quasi singolari, che rendono difficile la determinazione del rango. Per questo, si definisce il *rango numerico*:

Definizione 1.2. Se $\mathbf{A} \in \mathbb{R}^{m \times n}$, il *rango numerico* di \mathbf{A} è

$$rg(\mathbf{A}, \epsilon) := \min_{\|\mathbf{A}-\mathbf{B}\|_2 \leq \epsilon} rg(\mathbf{B}).$$

Si dice, quindi, che \mathbf{A} ha *rango numerico ridotto* (*numerically rank-deficient*) se

$$rg(\mathbf{A}, \epsilon) < \min\{m, n\},$$

con $\epsilon := u\|\mathbf{A}\|_2$ (u è la precisione della macchina).

Grazie alla SVD, possiamo trovare la distanza di una determinata matrice da una di rango inferiore:

Teorema 1.2.1. Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$, $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ la sua SVD.

Siano $k < r := \text{rg}(\mathbf{A})$ e

$$\mathbf{A}_k := \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

$$\text{Allora } \min_{\text{rg}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}.$$

Dimostrazione. $\mathbf{U}^T \mathbf{A}_k \mathbf{V} = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ \mathbf{0} & \dots & & \mathbf{0} \end{pmatrix}$ (per la definizione di \mathbf{A}_k),

quindi $\text{rg}(\mathbf{A}_k) = k$; inoltre, $\mathbf{U}^T(\mathbf{A} - \mathbf{A}_k)\mathbf{V} = \begin{pmatrix} \mathbf{0} & & & \\ & \sigma_{k+1} & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix}$,

quindi $\|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}$.

Sia ora $\mathbf{B} \in \mathbb{R}^{m \times n}$ tale che $\text{rg}(\mathbf{B}) = k$; allora possiamo trovare $\mathbf{x}_1, \dots, \mathbf{x}_{n-k}$ vettori ortonormali tali che $\ker(\mathbf{B}) = \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_{n-k}\}$. Vale

$$\text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_{n-k}\} \cap \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\} \neq \{0\} \text{ (ha dimensione 1),}$$

quindi esiste $\mathbf{z} \in \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_{n-k}\} \cap \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ tale che $\|\mathbf{z}\|_2 = 1$; allora avremo

$$\mathbf{B}\mathbf{z} = \mathbf{0} \text{ e } \mathbf{A}\mathbf{z} = \sum_{i=1}^{k+1} \sigma_i (\mathbf{v}_i^T \mathbf{z}) \mathbf{u}_i.$$

Quindi

$$\|\mathbf{A} - \mathbf{B}\|_2^2 \geq \|(\mathbf{A} - \mathbf{B})\mathbf{z}\|_2^2 = \|\mathbf{A}\mathbf{z}\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (\mathbf{v}_i^T \mathbf{z})^2 \geq \sigma_{k+1}^2.$$

□

Dal teorema segue che:

- il più piccolo valore singolare di \mathbf{A} è la distanza (in norma-2) di \mathbf{A} dall'insieme delle matrici a rango ridotto;
- l'insieme delle matrici in $\mathbb{R}^{m \times n}$ di rango pieno è aperto e denso.

Infine, se $r_\epsilon := rg(\mathbf{A}, \epsilon)$, allora

$$\sigma_1 \geq \dots \geq \sigma_{r_\epsilon} > \epsilon \geq \sigma_{r_\epsilon+1} \geq \dots \geq \sigma_p,$$

con $p = \min\{m, n\}$.

1.3 Aspetti computazionali della SVD

Esistono diversi metodi per calcolare effettivamente le matrici \mathbf{U}, \mathbf{V} e $\mathbf{\Sigma}$; la dimostrazione del Teorema 1.1.1 è costruttiva, e ne è un primo esempio. Altri metodi sfruttano la decomposizione in autovalori; ne mostreremo due in particolare.

Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$; la sua SVD sarà $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, $\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix}$ con $\mathbf{U}_1 \in \mathbb{R}^{m \times n}$ e definiamo

$$\mathbf{\Sigma}_1 := \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Il primo metodo riprende quanto visto nell'*Osservazione 3*:

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Sigma}_1^2 \mathbf{V}^T$$

è la decomposizione in autovalori di $\mathbf{A}^T \mathbf{A}$, con \mathbf{V} matrice ortonormale di autovettori e $\mathbf{\Sigma}_1^2$ diagonale con autovalori $\lambda_i \geq 0$ per $i = 1, \dots, n$; allora

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

Utilizzando la relazione $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$, otterremo quindi anche \mathbf{U} .

Gli autovalori non nulli di $\begin{pmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix}$ coincidono con i valori singolari di \mathbf{A} , comparando in coppie $\pm\sigma_i$; otterremo quindi maggiore accuratezza nel calcolo dei vettori singolari e dei valori singolari, in quanto i valori singolari non vengono elevati al quadrato, ma una dimensione molto maggiore della matrice.

Capitolo 2

La Decomposizione Polare

2.1 Esistenza della Decomposizione

La Decomposizione Polare [2] è una generalizzazione alle matrici della rappresentazione esponenziale dei numeri complessi: $z = re^{i\theta}$ con $r \geq 0$.

Teorema 2.1.1. *Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $m \geq n$. Allora esistono $\mathbf{U} \in \mathbb{R}^{m \times n}$ con colonne ortonormali e $\mathbf{H} \in \mathbb{R}^{n \times n}$ (univocamente determinata) simmetrica semidefinita positiva tali che*

$$\mathbf{A} = \mathbf{U}\mathbf{H}.$$

Se $\text{rg}(\mathbf{A}) = n$, allora \mathbf{H} è definita positiva e \mathbf{U} è univocamente determinata.

Dimostrazione. Sfruttando la Decomposizione ai Valori Singolari, avremo

$$\mathbf{A} = \tilde{\mathbf{U}}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}^T$$

con $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times m}$ ortogonale, $\tilde{\mathbf{U}} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix}$ con $\mathbf{U}_1 \in \mathbb{R}^{m \times n}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$

ortogonale e $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & \mathbf{0} & \end{pmatrix} \in \mathbb{R}^{m \times n}$. Allora

$$\mathbf{A} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}^T = \mathbf{U}_1\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}_1\mathbf{V}^T = (\mathbf{U}_1\mathbf{V}^T)(\mathbf{V}\mathbf{\Sigma}_1\mathbf{V}^T).$$

La tesi è verificata per $\mathbf{U} = \mathbf{U}_1 \mathbf{V}^T$ e $\mathbf{H} = \mathbf{V} \boldsymbol{\Sigma}_1 \mathbf{V}^T$. \square

Osservazione 4. Notiamo che, data la Decomposizione Polare di una matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$, si può risalire alla sua SVD ridotta: se $\mathbf{A} = \mathbf{U}\mathbf{H}$, facendo la decomposizione spettrale di \mathbf{H} , avremo

$$\mathbf{A} = \mathbf{U}\mathbf{H} = \mathbf{U}(\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T) = (\mathbf{U}\mathbf{Q})\boldsymbol{\Sigma}\mathbf{Q}^T$$

con \mathbf{Q} ortogonale.

2.2 Proprietà della Decomposizione

La Decomposizione Polare gode di alcune proprietà interessanti:

Proposizione 2.2.1. *Sia $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{A} = \mathbf{U}\mathbf{H}$ la sua Decomposizione Polare. Indichiamo con $\lambda(\mathbf{H})$ l'insieme degli autovalori di \mathbf{H} , $\sigma(\mathbf{H})$ l'insieme dei valori singolari di \mathbf{H} (analogo per \mathbf{A}) e k_2 il numero di condizionamento in norma-2. Allora:*

1. $\mathbf{H} = (\mathbf{A}^T \mathbf{A})^{\frac{1}{2}}$;
2. $\lambda(\mathbf{H}) = \sigma(\mathbf{H}) = \sigma(\mathbf{A})$;
3. $k_2(\mathbf{H}) = k_2(\mathbf{A})$;
4. $\mathbf{A}^T \mathbf{A} = \mathbf{A}\mathbf{A}^T$ (quindi \mathbf{A} normale) $\Leftrightarrow \mathbf{U}\mathbf{H} = \mathbf{H}\mathbf{U}$.

Dimostrazione. 1. $\mathbf{A}^T \mathbf{A} = \mathbf{H}\mathbf{U}^T \mathbf{U}\mathbf{H} = \mathbf{H}^2$, poiché \mathbf{U} ortogonale.

2. Per quanto visto nell'*Osservazione 3* del Capitolo 1,

$$\sigma_i(\mathbf{H}) = |\lambda_i(\mathbf{H})| = \lambda_i(\mathbf{H}), \text{ per } i = 1, \dots, n$$

dato che \mathbf{H} è simmetrica semidefinita positiva. Uso la decomposizione spettrale di $\mathbf{H} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ e la SVD di $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ con il punto 1.:

$$\mathbf{H}^2 = \mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^T = \mathbf{A}^T \mathbf{A} = \mathbf{V}\boldsymbol{\Sigma}^T \mathbf{U}^T \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \mathbf{V}\boldsymbol{\Sigma}_1^2 \mathbf{V}^T$$

quindi $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_1$.

3. Viene direttamente dal punto 2.
4. $\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T \Leftrightarrow \mathbf{H}^2 = \mathbf{U} \mathbf{H}^2 \mathbf{U}^T \Leftrightarrow \mathbf{H}^2 \mathbf{U} = \mathbf{U} \mathbf{H}^2 \Rightarrow \mathbf{H} \mathbf{U} = \mathbf{U} \mathbf{H}$.
Viceversa, se $\mathbf{H} \mathbf{U} = \mathbf{U} \mathbf{H} \Rightarrow \mathbf{H} \mathbf{U}^T = \mathbf{U}^T \mathbf{H}$; quindi

$$\mathbf{U} \mathbf{H}^2 \mathbf{U}^T = \mathbf{U} \mathbf{H} \mathbf{H} \mathbf{U}^T = \mathbf{H} \mathbf{U} \mathbf{U}^T \mathbf{H} = \mathbf{H}^2.$$

□

L'importanza della Decomposizione Polare sta inoltre nella proprietà di miglior approssimazione posseduta dal fattore ortogonale, come mostrato nel prossimo teorema:

Teorema 2.2.2. *Siano $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$ e $\mathbf{B}^T \mathbf{A} = \mathbf{U} \mathbf{H}$ la Decomposizione Polare di $\mathbf{B}^T \mathbf{A}$. Allora per ogni $\mathbf{Z} \in \mathbb{R}^{n \times n}$ si ha*

$$\|\mathbf{A} - \mathbf{B} \mathbf{U}\|_F \leq \|\mathbf{A} - \mathbf{B} \mathbf{Z}\|_F,$$

$$\text{con } \|\mathbf{A} - \mathbf{B} \mathbf{U}\|_F^2 = \sum_{i=1}^n (\sigma_i(\mathbf{A})^2 - 2\sigma_i(\mathbf{B}^T \mathbf{A}) + \sigma_i(\mathbf{B})^2).$$

Sfrutteremo questo risultato nel Capitolo 3.

Un caso particolare di questo teorema si ottiene considerando $m = n$ e $\mathbf{B} = \mathbf{I}$:

Corollario 2.2.3. *Sia $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{A} = \mathbf{U} \mathbf{H}$ la sua Decomposizione Polare. Allora per ogni $\mathbf{Z} \in \mathbb{R}^{n \times n}$*

$$\|\mathbf{A} - \mathbf{U}\|_F \leq \|\mathbf{A} - \mathbf{Z}\|_F,$$

$$\text{con } \|\mathbf{A} - \mathbf{U}\|_F = \left(\sum_{i=1}^n (\sigma_i(\mathbf{A}) - 1)^2 \right)^{1/2}.$$

Il corollario mostra quindi che la matrice ortogonale più vicina ad \mathbf{A} (in norma di Frobenius) è proprio il fattore ortogonale \mathbf{U} della Decomposizione Polare di \mathbf{A} .

Capitolo 3

L'analisi di Procuste

3.1 Che cos'è l'analisi di Procuste

L'analisi di Procuste [3] prende il nome dal brigante della mitologia greca classica Procuste, che aggrediva i viandanti e li sistemava su un'incudine a forma di letto scavata nella roccia, stirandoli a forza se troppo corti o amputandoli se troppo alti.

In statistica applicata, essa viene applicata a due matrici \mathbf{A} , $\mathbf{B} \in \mathbb{R}^{n \times p}$, rappresentanti ad esempio due configurazioni ottenute tramite metodi diversi di Scaling Multidimensionale¹ applicati a una serie di dati osservati sperimentalmente. L'obiettivo è confrontare numericamente le configurazioni, ruotando ed eventualmente traslando la seconda configurazione verso la prima.

Ogni riga di \mathbf{A} e \mathbf{B} , quindi, contiene le coordinate di uno degli n oggetti osservati; si tratta di trovare la matrice ortogonale \mathbf{Q} che minimizza

$$\|\mathbf{A} - \mathbf{BQ}^T\|_F^2.$$

Più in generale, si può pensare di ruotare e traslare la seconda configurazione in modo che si "avvicini" il più possibile alla prima; quindi, se

¹Lo Scaling Multidimensionale consiste nel trovare una configurazione di dati a dimensione ridotta rispetto all'originale che mantenga il più possibile le distanze osservate sperimentalmente tra ogni coppia di dati.

$\mathbf{c} \in \mathbb{R}^p$,

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{pmatrix},$$

vogliamo trovare

$$PR^2 := \min_{\mathbf{Q}, \mathbf{c}} \sum_{j=1}^n (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T - \mathbf{c})^T (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T - \mathbf{c}), \quad (1)$$

definita la *somma dei quadrati residui di Procuste*.

Il seguente teorema riprende il Teorema 2.2.2, calcolando esplicitamente PR^2 e mostrando una delle possibili applicazioni della SVD e della Decomposizione Polare:

Teorema 3.1.1. *Siano $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times p}$ centrate in modo che tutte le righe abbiano media $\mathbf{0}$ cioè*

$$\sum_{j=1}^n \mathbf{a}_j = \mathbf{0}, \quad \sum_{j=1}^n \mathbf{b}_j = \mathbf{0}.$$

Sia $\mathbf{B}^T \mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ la SVD di $\mathbf{B}^T \mathbf{A}$ con $\mathbf{U} = (\mathbf{u}_1 \dots \mathbf{u}_p)$, $\mathbf{V} = (\mathbf{v}_1 \dots \mathbf{v}_p)$ e

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{pmatrix}.$$

Sia $\mathbf{B}^T \mathbf{A} = \tilde{\mathbf{U}}\tilde{\mathbf{H}}$ la Decomposizione Polare di $\mathbf{B}^T \mathbf{A}$.

Allora

$$PR^2 = \sum_{j=1}^n \mathbf{a}_j \mathbf{a}_j^T + \sum_{j=1}^n \mathbf{b}_j \mathbf{b}_j^T - 2 \sum_{i=1}^p \sigma_i = \|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 - 2 \operatorname{tr}(\mathbf{\Sigma}),$$

e la coppia $(\hat{\mathbf{Q}}, \hat{\mathbf{c}})$ minimizzante PR^2 in (1) è data da

$$\hat{\mathbf{Q}} = \sum_{i=1}^p \mathbf{v}_i \mathbf{u}_i^T = \mathbf{V}\mathbf{U}^T = \tilde{\mathbf{U}}^T, \quad \hat{\mathbf{c}} = \mathbf{0}.$$

Dimostrazione. Dato che \mathbf{A} e \mathbf{B} sono centrate, avremo

$$\sum_{j=1}^n (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T - \mathbf{c})^T (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T - \mathbf{c}) = \sum_{j=1}^n (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T)^T (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T) + n \mathbf{c}^T \mathbf{c}.$$

L'ultimo termine è non negativo quindi la scelta migliore di \mathbf{c} è $\hat{\mathbf{c}} = \mathbf{0}$. Resta allora

$$\begin{aligned} PR^2 &= \min_{\mathbf{Q}} \sum_{j=1}^n (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T)^T (\mathbf{a}_j^T - \mathbf{Q}\mathbf{b}_j^T) = \\ &= \sum_{j=1}^n \mathbf{a}_j \mathbf{a}_j^T + \sum_{j=1}^n \mathbf{b}_j \mathbf{b}_j^T - 2 \max_{\mathbf{Q}} \sum_{j=1}^n \mathbf{a}_j \mathbf{Q}\mathbf{b}_j^T. \end{aligned}$$

Usando la relazione $\mathbf{a}_j \mathbf{Q}\mathbf{b}_j^T = \text{tr}(\mathbf{Q}\mathbf{b}_j^T \mathbf{a}_j)$, l'espressione da massimizzare diventa

$$\sum_{j=1}^n \mathbf{a}_j \mathbf{Q}\mathbf{b}_j^T = \sum_{j=1}^n \text{tr}(\mathbf{Q}\mathbf{b}_j^T \mathbf{a}_j) = \text{tr}(\mathbf{Q} \sum_{j=1}^n \mathbf{b}_j^T \mathbf{a}_j).$$

Per la SVD,

$$\sum_{j=1}^n \mathbf{b}_j^T \mathbf{a}_j = \mathbf{B}^T \mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T;$$

quindi

$$\sum_{j=1}^n \mathbf{a}_j \mathbf{Q}\mathbf{b}_j^T = \text{tr}(\mathbf{Q} \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T) = \sum_{i=1}^p \sigma_i \text{tr}(\mathbf{Q}\mathbf{u}_i \mathbf{v}_i^T).$$

Per la disuguaglianza di Cauchy-Schwarz applicata a $\mathbf{Q}\mathbf{v}_i$ e \mathbf{u}_i , dato che \mathbf{U} , \mathbf{V} e \mathbf{Q} sono ortogonali,

$$\text{tr}(\mathbf{Q}\mathbf{u}_i \mathbf{v}_i^T) = \mathbf{v}_i^T \mathbf{Q}\mathbf{u}_i \leq 1$$

e, scegliendo $\hat{\mathbf{Q}} = \mathbf{V}\mathbf{U}^T$,

$$\mathbf{v}_i^T \hat{\mathbf{Q}} \mathbf{u}_i = \mathbf{v}_i^T \mathbf{V}\mathbf{U}^T \mathbf{u}_i = (\mathbf{0} \ 1 \ \mathbf{0}) \begin{pmatrix} \mathbf{0} \\ 1 \\ \mathbf{0} \end{pmatrix} = 1.$$

Quindi

$$-2 \max_{\mathbf{Q}} \sum_{j=1}^n \mathbf{a}_j \mathbf{Q}\mathbf{b}_j^T = -2(\sigma_1 + \dots + \sigma_p).$$

In ultimo, $\hat{\mathbf{Q}}\hat{\mathbf{Q}}^T = \mathbf{V}\mathbf{U}^T\mathbf{U}\mathbf{V}^T = I_p$ quindi $\hat{\mathbf{Q}}$ è ortogonale. □

3.2 Un esempio di analisi di Procuste

La Tabella 3.1 [3] riporta dati relativi a 25 università americane, e comprende il punteggio medio dei SAT dei nuovi studenti, la percentuale di nuovi studenti che si trovavano nella top 10% delle classi di scuola superiore, la percentuale di domande accettate, il rapporto studenti-facoltà, la spesa annuale stimata e la percentuale di laureati.

University	SAT	Top10	Accept	SFRatio	Expenses	Grad
Harvard	14.00	91	14	11	39.525	97
Princeton	13.75	91	14	8	30.220	95
Yale	13.75	95	19	11	43.514	96
Stanford	13.60	90	20	12	36.450	93
MIT	13.80	94	30	10	34.870	91
Duke	13.15	90	30	12	31.585	95
CalTech	14.15	100	25	6	63.575	81
Dartmouth	13.40	89	23	10	32.162	95
Brown	13.10	89	22	13	22.704	94
JohnsHopkins	13.05	75	44	7	58.691	87
UChicago	12.90	75	50	13	38.380	87
UPenn	12.85	80	36	11	27.553	90
Cornell	12.80	83	33	13	21.864	90
Northwestern	12.60	85	39	11	28.052	89
Columbia	13.10	76	24	12	31.510	88
NotreDame	12.55	81	42	13	15.122	94
UVir	12.25	77	44	14	13.349	92
Georgetown	12.55	74	24	12	20.126	92
CarnegieMellon	12.60	62	59	9	25.026	72
UMichigan	11.80	65	68	16	15.470	85
UCBerkeley	12.40	95	40	17	15.140	78
UWisconsin	10.85	40	69	15	11.857	71
PennState	10.81	38	54	18	10.185	80
Purdue	10.05	28	90	19	9.066	69
TexasA&M	10.75	49	67	25	8.704	67

Tabella 3.1: Dati relativi alle università.

Da due diversi metodi di Scaling Multidimensionale otteniamo due configurazioni bidimensionali dei dati, raccolti in matrici 25×2 , in cui ogni riga si riferisce a una differente università:

$$\mathbf{A} = \begin{pmatrix} -2.1951 & -0.3436 \\ -2.0321 & -0.6833 \\ -2.1018 & -0.1181 \\ -1.5808 & -0.2507 \\ -1.5773 & 0.0812 \\ -1.1484 & -0.3870 \\ -3.0249 & 1.8984 \\ -1.4897 & -0.3941 \\ -0.9216 & -0.8092 \\ -1.4687 & 1.9439 \\ -0.1692 & 0.6474 \\ -0.5303 & -0.1355 \\ -0.3395 & -0.4616 \\ -0.4296 & -0.0579 \\ -0.8218 & -0.0308 \\ 0.0773 & -0.7258 \\ 0.4850 & -0.6392 \\ -0.3894 & -0.7441 \\ 1.0437 & 1.7378 \\ 1.8268 & -0.0874 \\ 0.9053 & -1.4828 \\ 3.2912 & 1.0151 \\ 3.1574 & -0.0040 \\ 4.8761 & 0.8998 \\ 4.5574 & -0.8687 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -2.3698 & -0.3907 \\ -2.2539 & -0.3778 \\ -2.3030 & -0.1241 \\ -1.6971 & -0.3143 \\ -1.7225 & 0.0882 \\ -1.2448 & -0.4760 \\ -2.8175 & 2.2872 \\ -1.6805 & -0.3096 \\ -1.0033 & -1.0301 \\ -1.4049 & 2.1980 \\ -0.1123 & 0.8007 \\ -0.5505 & -0.0366 \\ -0.3186 & -0.6046 \\ -0.4664 & 0.0207 \\ -0.7650 & -0.0145 \\ 0.0283 & -1.0921 \\ 0.5222 & -1.0547 \\ -0.3583 & -0.8793 \\ 1.1043 & 1.8999 \\ 2.0065 & -0.1945 \\ 0.8805 & -1.4554 \\ 3.4951 & 0.8955 \\ 3.2678 & -0.2981 \\ 5.0683 & 0.8900 \\ 4.6954 & -0.4279 \end{pmatrix}$$

La somma iniziale dei quadrati delle distanze è

$$\sum_{j=1}^{25} \|\mathbf{a}_j - \mathbf{b}_j\|_2^2 = \|\mathbf{A} - \mathbf{B}\|_F^2 = 1.5122$$

e il seguente grafico mostra entrambe le configurazioni (Figura 3.1):

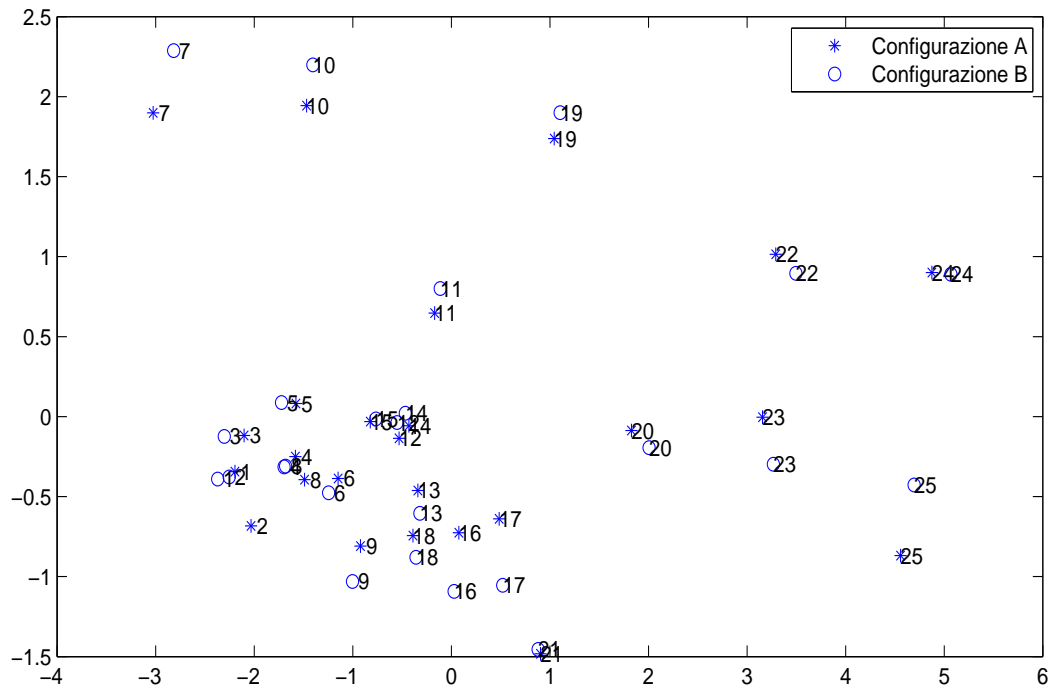


Figura 3.1: Rappresentazioni bidimensionali dei dati prodotte da due metodi diversi di Scaling Multidimensionale

Calcoliamo poi la SVD di $\mathbf{B}^T \mathbf{A}$:

$$\mathbf{U} = \begin{pmatrix} -0.9999 & 0.0107 \\ 0.0107 & 0.9999 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} 110.6718 & 0 \\ 0 & 20.7442 \end{pmatrix},$$

$$\mathbf{V} = \begin{pmatrix} -0.9999 & -0.0102 \\ -0.0102 & 0.9999 \end{pmatrix}.$$

Dal Teorema 3.1.1, sappiamo che la trasformazione che minimizza $\|\mathbf{A} - \mathbf{B}\hat{\mathbf{Q}}^T\|$ è

$$\hat{\mathbf{Q}} = \mathbf{V}\mathbf{U}^T = \begin{pmatrix} 0.9998 & -0.0210 \\ 0.0210 & 0.9998 \end{pmatrix}.$$

Dopo la rotazione, quindi, la matrice \mathbf{B} diventerà

$$\mathbf{B}\hat{\mathbf{Q}}^T = \begin{pmatrix} -2.3611 & -0.4403 \\ -2.2454 & -0.4250 \\ -2.2999 & -0.1724 \\ -1.6902 & -0.3498 \\ -1.7240 & 0.0521 \\ -1.2345 & -0.5019 \\ -2.8648 & 2.2277 \\ -1.6736 & -0.3448 \\ -0.9815 & -1.0510 \\ -1.4507 & 2.1681 \\ -0.1291 & 0.7981 \\ -0.5496 & -0.0481 \\ -0.3058 & -0.6111 \\ -0.4667 & 0.0109 \\ -0.7646 & -0.0305 \\ 0.0512 & -1.0913 \\ 0.5442 & -1.0435 \\ -0.3398 & -0.8866 \\ 1.0642 & 1.9227 \\ 2.0101 & -0.1524 \\ 0.9109 & -1.4366 \\ 3.4755 & 0.9686 \\ 3.2733 & -0.2295 \\ 5.0486 & 0.9961 \\ 4.7034 & -0.3294 \end{pmatrix} .$$

Il seguente grafico mostra le configurazioni originali e la configurazione ottenuta:

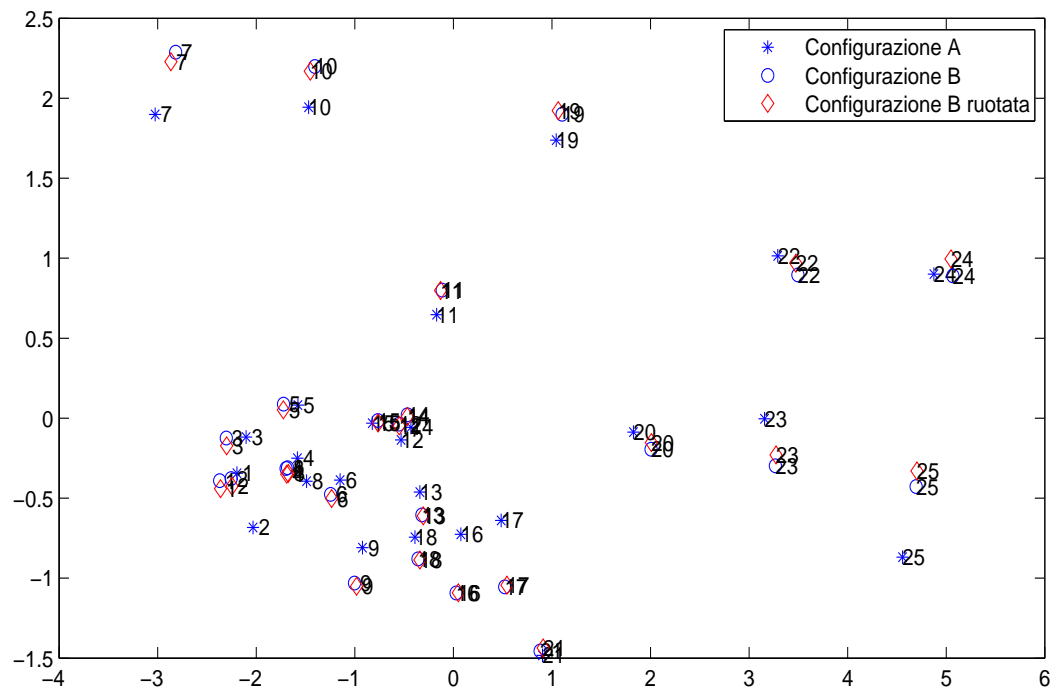


Figura 3.2: Rappresentazione delle due configurazioni iniziali e della nuova configurazione ruotata.

La somma dei quadrati residui di Procuste sarà infine

$$PR^2 = \|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 - 2 \operatorname{tr}(\mathbf{\Sigma}) = 1.4545.$$

Appendice A

L'algoritmo MATLAB

Per calcolare le varie configurazioni e ottenere i grafici, è stato usato il seguente algoritmo MATLAB:

```
load t12-9-x.txt
Tscaled=zscore(t12_9_x);      % standardizzazione dei dati
[n,p]=size(Tscaled);
D=pdist(Tscaled);           % calcolo delle distanze tra coppie di dati

% scaling multidimensionale non metrico
[A,stress1]=mdscale(D,2);
plot(A(:,1),A(:,2),'*')
text(A(:,1),A(:,2),num2str((1:n)'))
hold on

% scaling multidimensionale metrico
[B,stress2]=mdscale(D,2,'criterion','metricsstress');
plot(B(:,1),B(:,2),'o')
text(B(:,1),B(:,2),num2str((1:n)'))

% somma iniziale dei quadrati delle distanze
isq=norm(A-B,'fro')^2
```

```
% calcolo SVD
[uu,ss,vv]=svd(B'*A);

% trasformazione minimizzante
Q=vv*uu';

% somma dei quadrati residui di Procuste
PR2=norm(A-B*Q','fro')^2

% nuova configurazione
Bp=(B*Q');
plot(Bp(:,1),Bp(:,2),'rd')
text(Bp(:,1),Bp(:,2),num2str((1:n)'))
```

Bibliografia

- [1] *Matrix Computations*, G. H. Golub e C. F. Van Loan, The Johns Hopkins University Press, II ed. 1989.
- [2] *Computing the Polar Decomposition - with Applications*, N. J. Higham, SIAM, 1986.
- [3] *Multivariate Statistical Analysis*, R. A. Johnson e D. W. Winchern, Prentice Hall, V ed. 2002.