

SVD

Federico Riva
Ingegneria Matematica
Politecnico di Milano

Relatore: Professor Marco Verani

Indice

1 Decomposizione SVD

1.1 Introduzione

La decomposizione SVD è una fattorizzazione algebrica che consente di rappresentare una matrice come il prodotto di tre matrici ciascuna con caratteristiche specifiche. Le applicazioni di questo metodo possono essere ritrovate in una vasta gamma di campi tra cui l'analisi dei dati, il machine learning e l'elaborazione delle immagini.

1.2 Premessa Teorica

Prima di entrare nel vivo della trattazione della fattorizzazione SVD della matrici è necessario fornire una serie di definizioni, Teoremi e proposizioni preliminari utili per la comprensione dei passaggi che seguiranno.

Definizione 1. *Matrice Ortogonale*

Una matrice quadrata $Q \in \mathbb{R}^{n \times n}$ si dice ortogonale se $Q^\top Q = I$.

Definizione 2. *Matrice Pseudodiagonale*

Una matrice M si dice pseudodiagonale se $m_{ij} = 0 \ \forall i \neq j$.

Definizione 3. *Matrice Simmetrica*

Una matrice quadrata $Q \in \mathbb{R}^{n \times n}$ si dice simmetrica se $Q^\top = Q$.

Definizione 4. *Forma Quadratica*

Si definisce forma quadratica un polinomio di secondo grado in n variabili $p: \mathbb{R}^n \rightarrow \mathbb{R}$ t.c.

$$p(x) = p(x_1, \dots, x_n) = \sum_{1 \leq i, j \leq n} a_{ij} x_i x_j = a_{11} x_1^2 + a_{12} x_1 x_2 + \dots + a_{nn-1} x_n x_{n-1} + a_{nn} x_n^2 = \mathbf{x}^\top A \mathbf{x}$$

dove $\mathbf{x} = (x_1, \dots, x_n)$ e A è una matrice simmetrica di dimensione $n \times n$.

Proposizione 1.

Hp: Data una qualsiasi $M \in \mathbb{R}^{m \times n}$

Ts: $D = M^\top M$ e $S = M M^\top$ sono matrici simmetriche e definite positive.

Dimostrazione. 1) Vogliamo dimostrare che $D = D^\top$. Segue quindi dalle proprietà di trasposizione del prodotto di due matrici: $D^\top = (M^\top M)^\top = M^\top (M^\top)^\top = M^\top M = D$. Si dimostra in maniera analoga il risultato per S .

2) Dal momento che è una matrice quadrata per dimostrare la semipositività della matrice $A^\top A$ possiamo studiare il segno della forma quadratica associata. Se quest'ultima sarà positiva allora lo saranno anche tutti gli autovalori.

$$\forall \mathbf{x} \neq 0 : \mathbf{x} \in \mathbb{R}^n$$
$$\mathbf{x}^\top A^\top A \mathbf{x} = (A \mathbf{x})^\top (A \mathbf{x}) = \|A \mathbf{x}\|^2 \geq 0$$

□

Osservazione: Poiché gli autovalori che otteniamo sono ≥ 0 possiamo calcolarne la radice quadrata.

Proposizione 2.

Hp: Sia $A \in \mathbb{R}^{m \times n}$ con $m \geq n$ con $\{\lambda_i(A^\top A)\}_{i=1}^n$ e $\{\lambda_j(AA^\top)\}_{j=1}^m$

Ts: $\{\lambda_i(A^\top A)\}_{i=1}^n \subset \{\lambda_j(AA^\top)\}_{j=1}^m$.

Dimostrazione. Ipotizziamo che la coppia autovalori autovettore (\mathbf{x}, λ) sia associata alla matrice $A^\top A$. Possiamo quindi scrivere

$$A^\top A \mathbf{x} = \lambda \mathbf{x}$$

e moltiplicando ambo i membri per A otteniamo

$$AA^\top A \mathbf{x} = \lambda A \mathbf{x}$$

Ma allora definendo $\mathbf{y} = A \mathbf{x}$ possiamo riscrivere

$$AA^\top \mathbf{y} = \lambda \mathbf{y}$$

Il che conclude la dimostrazione poiché prova che λ è un autovalore anche della matrice AA^\top . \square

Corollario 1.

Hp: Sia $M \in \mathbb{R}^{n \times n}$ con autovalori $\lambda = [\lambda_1, \dots, \lambda_n]$

Ts: Gli autovalori di $M^\top M$ e MM^\top sono gli stessi e valgono $\lambda^2 = [\lambda_1^2, \dots, \lambda_n^2]$

Dimostrazione. **Proposizione ??** nel caso particolare $A = M$ \square

Teorema 1. Teorema Spettrale

Hp: Data $Q \in \mathbb{R}^{n \times n}$ una matrice simmetrica

Ts: È ortogonalmente diagonalizzabile $Q = U^{-1}DU = U^\top DU$

Prendendo in considerazione le **Proposizioni ??**, **??** ed il **Teorema ??** possiamo osservare che data una qualsiasi matrice A è possibile costruire a partire da essa due matrici simmetriche ortogonalmente diagonalizzabili. Possiamo quindi fornire le seguenti definizioni.

Definizione 5. Vettori Singolari di sinistra

I vettori singolari di sinistra di A sono gli autovettori $\mathbf{u}_1, \dots, \mathbf{u}_m$ di $AA^\top = U^\top D_l U$ e compongono la Matrice Singolare di Sinistra U .

Definizione 6. Vettori Singolari di destra

I vettori singolari di destra di A sono gli autovettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ di $A^\top A = V^\top D_r V$ e compongono la Matrice Singolare di destra V .

Definizione 7. Valori Singolari

I valori singolari sono le radici quadrate degli autovalori non nulli di $A^\top A$ e $A^\top A$

$\sigma_1 \geq \dots \geq \sigma_{r=\min n, m} \geq 0$. Indicheremo con r il numero di valori singolari che nel caso di assenza di autovalori nulli sarà $r = \min \{m, n\}$.

1.3 Definizione

Sia $A \in \mathbb{R}^{m \times n}$, oppure equivalentemente $A \in \mathbb{C}^{m \times n}$, allora la decomposizione SVD della una matrice nel caso $m > n$ è definita come:

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

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \cdot \mathbf{v}_i^\top$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]^\top$$

Mentre per $m < n$:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \end{bmatrix} [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]^\top$$

Dove:

- $U \in \mathbb{R}^{m \times m}$ è una matrice ortogonale contenente i vettori singolari sinistri di A ,
- $\Sigma \in \mathbb{R}^{m \times n}$ è una matrice pseudodiagonale contenente i valori singolari di A ordinati in modo decrescente,
- $V \in \mathbb{R}^{n \times n}$ è una matrice ortogonale contenente i vettori singolari destri di A .

1.4 Legame fra la decomposizione SVD ed il Teorema Spettrale

Da un punto di vista intuitivo la fattorizzazione SVD può essere considerata come la generalizzazione del concetto di diagonalizzazione di una matrice quadrata; infatti se la seconda può essere applicata solamente a matrici quadrate ($n \times n$) risulta valida per qualsiasi tipo di matrice ($m \times n$) senza limite alcuno sulla sua forma.

Se A ammette la fattorizzazione SVD allora $A = U \Sigma V^\top$ valgono le seguenti relazioni:

$$A^\top A = (U \Sigma V^\top)^\top U \Sigma V^\top = V \Sigma^\top U^\top U \Sigma V^\top = V \Sigma^\top \Sigma V^\top$$

$$AA^\top = U\Sigma V^\top (U\Sigma V^\top)^\top = \dots = U\Sigma\Sigma^\top U^\top$$

Vale dunque un risultato ancora più specifico della semplice diagonalizzazione di una matrice ovvero il Teorema Spettrale

1.5 Interpretazione geometrica

Data una matrice $A \in \mathbb{R}^{m \times n}$ consideriamo l'applicazione lineare ad essa associata

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

La fattorizzazione SVD scompone la trasformazione lineare in 3 trasformazioni geometriche descritte dalle 3 matrici U, V e Σ dove le matrici ortonormali rappresentano una rotazione o una simmetria assiale, in generale una isometria, mentre la matrice Σ è una dilatazione di intensità σ_i lungo la componente \mathbf{x}_i . Di seguito illustriamo visivamente lo specifico caso $A \in \mathbb{R}^{2 \times 2}$ nel caso di una circonferenza di raggio unitario che viene mappata in un'ellisse.

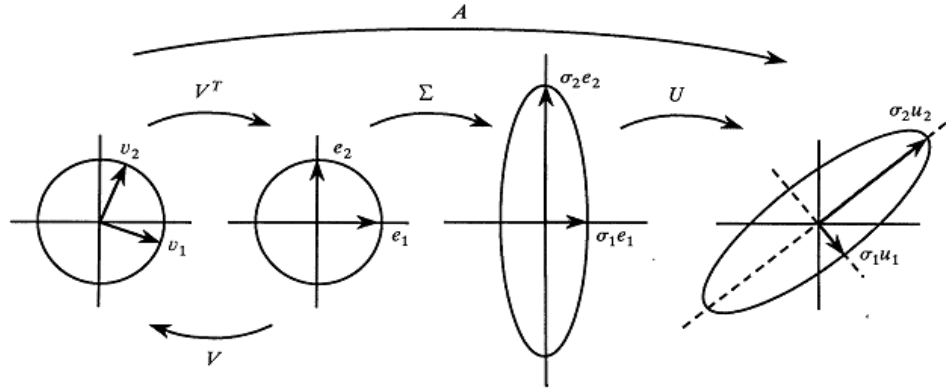


Figura 1: Illustrazione in \mathbb{R}^2

Possiamo estendere questo ragionamento ad un caso più generale considerando una sfera di raggio unitario in \mathbb{R}^n e l'ellissoide ad essa associata da $A \in \mathbb{R}^{m \times n}$ in \mathbb{R}^m .

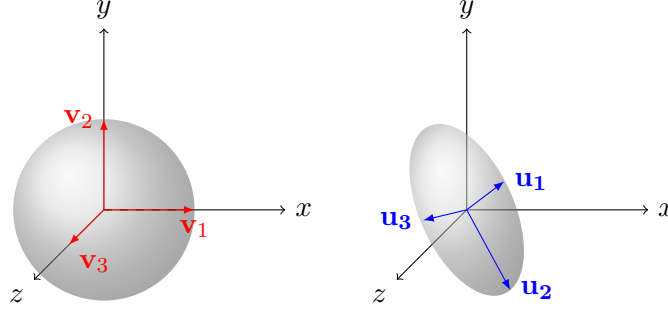


Figura 2: Mappa da una sfera in \mathbb{R}^n ad un'ellissoide in \mathbb{R}^m nel caso particolare $n, m = 3$.

Notiamo che le matrici U e V sono unitarie e le colonne di ciascuna formano una base di vettori ortonormali. Inoltre la matrice A mappa il v_i elemento della base in nel vettore dilatato $\sigma_i u_i$. Verifichiamolo:

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

$$AV = U\Sigma$$

$$[Av_1 \ Av_2 \ \dots \ Av_n] = [U\sigma_1 \ U\sigma_2 \ \dots \ U\sigma_n]$$

Tenendo conto che

$$\sigma_i = [0 \ \dots \ 0 \ \sigma_i \ 0 \ \dots \ 0]$$

Otteniamo

$$\forall i = 1, \dots, n \quad Av_i = \sigma_i u_i$$

Possiamo quindi scrivere più in generale che: $A\mathbf{v} = \sigma\mathbf{u}$ e $A^\top \mathbf{u} = \sigma\mathbf{v}$.

Studiamo ora le relazioni che sussistono fra gli spazi vettoriali associati alle colonne della matrice A e delle matrici di decomposizione U e V . Per farlo avremo bisogno di qualche risultato teorico pregresso.

Osservazione: $\text{Span}(A) = \text{Col}(A)$ e quindi $\text{Span}(A^\top) = \text{Col}(A^\top) = \text{Row}(A)$.

Proposizione 3.

Hp: Per ogni matrice $A \in \mathbb{R}^{m \times n}$ valgono le seguenti relazioni

Ts: $\text{Ker}(A) = \text{Row}(A)^\perp$ $\text{Row}(A) = \text{Ker}(A)^\perp$ e $\text{Ker}(A^\top) = \text{Col}(A)^\perp$ $\text{Col}(A) = \text{Ker}(A^\top)^\perp$

Teorema 2. Teorema di Nullità più rango

Hp: Sia $A \in \mathbb{R}^{m \times n}$ con $\text{Rg}(A) = r$

Ts: $\dim \text{Ker}(A) = n - r$

Proposizione 4.

Hp: $\forall A \in \mathbb{R}^{m \times n}$ valgono le seguenti uguaglianze

Ts: $\text{Ker}(A) = \text{Ker}(A^\top A)$, $\text{Col}(A^\top) = \text{Col}(A^\top A)$ e di conseguenza $\text{Rg}(A) = \text{Rg}(A^\top A)$

Dimostrazione. Osserviamo che

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^\top (A\mathbf{x}) = \mathbf{x}^\top A^\top A \mathbf{x}$$

Ne segue che, se \mathbf{x} appartiene al Nucleo di $A^\top A$ allora $\|A\mathbf{x}\|^2 = 0$ e quindi $A\mathbf{x} = 0$, cioè $\mathbf{x} \in \text{Ker}(A)$. Viceversa se $\mathbf{x} \in \text{Ker}(A)$ ripercorrendo la catena di uguaglianze a retroso $\mathbf{x} \in \text{Ker}(A^\top A)$ e per arbitrarietà della scelta del vettore i due nuclei sono uguali. Dall'uguaglianza dei nuclei per il **Teorema ??** segue l'uguaglianza dei ranghi.

$\text{Col}(A^\top) = \text{Row}(A)$ e $(\text{Row}(A))^\perp = \text{Ker}(A)$. Ma $\text{Ker}(A) = \text{Ker}(A^\top A)$ il cui complemento ortogonale è $\text{Row}(A^\top A)$. Quindi $\text{Col}(A^\top) = \text{Row}(A^\top A)$ ma per simmetria di $A^\top A$

$$\text{Row}(A^\top A) = \text{Col}(A^\top A) \Rightarrow \text{Col}(A) = \text{Col}(A^\top A)$$

□

Corollario 2.

$$\text{Ker}(A^\top) = \text{Ker}(AA^\top), \quad \text{Col}(A) = \text{Col}(AA^\top), \quad \text{Rg}(A^\top) = \text{Rg}(AA^\top).$$

Se quindi la matrice A non ha rango massimo, $\text{Rg}(A) = r < \min\{m, n\}$, allora sussiste la seguente relazione:

Proposizione 5.

Hp: Data una matrice $A \in \mathbb{R}^{m \times n}$ e la sua fattorizzazione SVD $A = U\Sigma V^\top$

Ts: Valgono le seguenti relazioni:

1. Le prime r colonne di U sono una base dello $\text{Span}(A)$;
2. Le ultime $m - r$ colonne di U sono una base del $\text{Ker}(A^\top)$;
3. Le prime r colonne di V sono una base dello $\text{Span}(A^\top)$;
4. Le ultime $n - r$ colonne di V sono una base del $\text{Ker}(A)$

Dimostrazione. 1. e 2.

$$\text{Col}(A) = \text{Col}(AA^\top)$$

$$\text{Col}(U\Sigma V^\top) = \text{Col}((U\Sigma V^\top)(U\Sigma V^\top)^\top) = \text{Col}(U\Sigma V^\top V \Sigma^\top U^\top) = \text{Col}(U\Sigma \Sigma^\top U^\top)$$

Sapendo che $\Sigma \in \mathbb{R}^{m \times n}$ e $\text{Rg}(\Sigma) = r$

$$\Sigma \Sigma^\top = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

quando svolgiamo il prodotto fra $U \Sigma \Sigma^\top$ e U^\top otteniamo

$$[\mathbf{u}_1 \sigma_1^2 \dots \mathbf{u}_r \sigma_r^2 0 \dots 0]$$

da cui la tesi $\text{Col}(A) = \text{Span}([\mathbf{u}_1 \dots \mathbf{u}_r])$

$$\text{Ker}(A^\top) = \text{Ker}(AA^\top)$$

$$\text{Ker}(V \Sigma^\top U^\top) = \text{Ker}((U \Sigma V^\top)(U \Sigma V^\top)^\top) = \text{Ker}(U \Sigma \Sigma^\top U)$$

e quindi $\text{Ker}(A^\top) = \text{Span}([\mathbf{u}_{m-r} \dots \mathbf{u}_m])$ \square

Dimostrazione. 3. e 4. Analogamente a 1. e 2. sfruttando le proprietà della trasposizione del prodotto matriciale \square

1.6 Dimostrazione della decomposizione SVD e Teorema Eckart-Young

Teorema 3. *Decomposizione a valori singolari*

Hp: Sia $A \in \mathbb{R}^{m \times n}$ e $\text{Rg}(A) = r$.

Ts: Esistono una matrice pseudodiagonale $\Sigma \in \mathbb{R}^{m \times n}$ con $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ e due matrici ortogonali U e V tali che possiamo riscrivere la matrice A come $A = U \Sigma V^\top$. Tale decomposizione si dice decomposizione ai valori singolari di A , non è unica, ma in ogni decomposizione siffatta gli elementi non nulli σ_k di Σ sono i valori singolari di A .

1.6.1 Dimostrazione

Data una matrice $A \in \mathbb{R}^{m \times n}$ poiché $A^\top A \in \mathbb{R}^{n \times n}$ è simmetrica e semidefinita positiva segue dal Teorema Spettrale che \exists una matrice ortonormale $V \in \mathbb{R}^{n \times n}$ tale che

$$A^\top A = V \bar{D} V^\top$$

e a sua volta poiché $V^{-1} = V^\top$

$$V^\top A^\top A V = \bar{D} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

dove $D \in \mathbb{R}^{\ell \times \ell}$ è una matrice diagonale e positiva di dimensione con $\ell \leq \min\{m, n\}$ il numero di autovalori non nulli della matrice $A^\top A$. Osserviamo che per definizione di autovalori e autovettori la i -esima colonna di V corrisponde all' i -esimo autovalore \bar{D}_{ii} . Di conseguenza possiamo individuare $V_1 = [\mathbf{v}_1 \dots \mathbf{v}_\ell] \in \mathbb{R}^{n \times \ell}$ e $V_2 = [\mathbf{v}_{\ell+1} \dots \mathbf{v}_n] \in \mathbb{R}^{n \times n-\ell}$ tali che $V = [V_1 \ V_2]$ che sono rispettivamente gli autovettori associati ad autovalori non nulli e nulli. Possiamo quindi riscrivere l'equazione come:

$$\begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix} A^\top A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} V_1^\top A^\top A V_1 & V_1^\top A^\top A V_2 \\ V_2^\top A^\top A V_1 & V_2^\top A^\top A V_2 \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

Da cui ricaviamo le seguenti equazioni:

$$V_1^\top A^\top A V_1 = D \quad \text{e} \quad V_2^\top A^\top A V_2 = 0$$

Siamo quindi in grado di risolvere la seconda

$$\|AV_2\|_2 = 0 \Rightarrow AV_2 = 0$$

Costruiamo inoltre le seguenti matrici di diverse dimensioni che torneranno utili in seguito:

- $V_1^\top V_1 = I_\ell$ dove $(\ell \times n) \times (n \times \ell) = (\ell \times \ell)$
- $V_2^\top V_2 = I_{n-\ell}$ dove $(n-\ell \times n) \times (n \times n-\ell) = (n-\ell \times n-\ell)$
- $[V_1 \ V_2][V_1 \ V_2]^\top = V_1 V_1^\top + V_2 V_2^\top = I_n$ dove

$$(n \times \ell) \times (\ell \times n) = (n \times n) \quad \text{e} \quad (n \times n-\ell) \times (n-\ell \times n) = (n \times n)$$

Definiamo ora la matrice U_1 come:

$$U_1 = AV_1 D^{-\frac{1}{2}} \in \mathbb{R}^{m \times \ell} \{(m \times \ell) = (m \times n) \times (n \times \ell) \times (\ell \times \ell)\}$$

sapendo che $D^{-\frac{1}{2}}$ è ben definita poiché $D_{ii} > 0 \ \forall i = 1, \dots, \ell$.

Invertendo la relazione e sostituendo U_1 :

$$A = U_1 D^{\frac{1}{2}} V_1^\top = AV_1 D^{-\frac{1}{2}} D^{\frac{1}{2}} V_1^\top = AV_1 V_1^\top$$

e sfruttando $V_1 V_1^\top + V_2 V_2^\top = I_n$ e $AV_2 = 0$ otteniamo:

$$A = A \left(I - V_2 V_2^\top \right) = A - AV_2 V_2^\top = A - 0 = A$$

il che stabilisce la coerenza dalla definizione di U_1 proposta.

Abbiamo quindi quasi dimostrato il risultato completo, non ci resta che verificare che le colonne della matrice U_1 formino una base ortonormale:

$$U_1^\top U_1 = D^{-\frac{1}{2}} V_1^\top A^\top AV_1 D^{-\frac{1}{2}} = D^{-\frac{1}{2}} D D^{-\frac{1}{2}} = I_\ell \in \mathbb{R}^{\ell \times \ell}$$

Possiamo quindi costruire il complemento ortogonale di U_1 ovvero $U_2 = \text{Span}(\text{Col}(U_1)^\perp) \in \mathbb{R}^{m \times n - \ell}$ in modo che $U = [U_1 \ U_2] \in \mathbb{R}^{m \times m}$ sia una matrice ortogonale.

Per concludere aggiungiamo un numero opportuno di righe e colonne nulle alla matrice $D^{\frac{1}{2}} \in \mathbb{R}^{\ell \times \ell}$ in modo da creare:

$$\Sigma = \begin{bmatrix} \begin{bmatrix} D^{\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} \\ 0 \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Allora sfruttando le proprietà del prodotto matriciale fra matrici a blocchi:

$$U\Sigma V^\top = [U_1 \ U_2] \begin{bmatrix} \begin{bmatrix} D^{\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} \\ 0 \end{bmatrix} [V_1 \ V_2]^\top = [U_1 \ U_2] \begin{bmatrix} D^{\frac{1}{2}} V_1^\top \\ 0 \end{bmatrix} = U_1 D^{\frac{1}{2}} V_1^\top = A$$

□

1.6.2 Norme Matriciali e Proprietà

Introduciamo ora alcune nozioni sulle norme matriciali che adopereremo per discutere alcuni risultati di approssimazione. Dal momento che $\mathbb{R}^{m \times n}$ è isomorfo a \mathbb{R}^{mn} la norma di una matrice è equivalente alla norma di un vettore.

Definizione 8. Norma

Una funzione $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ è detta norma se soddisfa le seguenti proprietà

- $f(A) \geq 0 \ \forall A \in \mathbb{R}^{m \times n}$ e $f(A) = 0 \iff A = 0$
- $f(A + B) \leq f(A) + f(B) \ \forall A, B \in \mathbb{R}^{m \times n}$
- $f(\alpha A) = |\alpha|f(A), \ \alpha \in \mathbb{R}, A \in \mathbb{R}^{m \times n}$

Useremo quindi la notazione $\|A\| = f(A)$ per indicare la 'norma della matrice A '.

Fra le numerose norme esistenti le due più frequenti sono:

Definizione 9. Norma di Frobenius

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

Definizione 10. Norma p-esima

$$\|A\|_p = \sup_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p$$

Definizione 11. Norma ∞

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$$

Proposizione 6.

Hp: Data una matrice ortogonale $Q \in \mathbb{R}^{n \times n}$

Ts: $\forall \mathbf{x} \in \mathbb{R}^n \quad \|Q\mathbf{x}\| = \|\mathbf{x}\|$

Dimostrazione. $\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^\top (Q\mathbf{x}) = \mathbf{x}^\top Q^\top Q \mathbf{x} = \mathbf{x}^\top I \mathbf{x} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2$ □

Enunciamo ora una serie di corollari che seguono dalla dimostrazione della fattorizzazione SVD.

Corollario 3.

Hp: Sia $U^\top A V = \Sigma$ la decomposizione SVD della matrice $A \in \mathbb{R}^{m \times n}$ con $m \geq n$.

Ts: $\forall i = 1, \dots, n \quad A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ e $A^\top \mathbf{u}_i = \sigma_i \mathbf{v}_i$

Dimostrazione. Come abbiamo già discusso precedentemente nell'interpretazione geometrica:

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

$$A V = U \Sigma$$

$$[A\mathbf{v}_1 \ A\mathbf{v}_2 \ \dots \ A\mathbf{v}_n] = [U\sigma_1 \ U\sigma_2 \ \dots \ U\sigma_n]$$

Tenendo conto che

$$\sigma_i = [0 \ \dots \ 0 \ \sigma_i \ 0 \ \dots \ 0]$$

Otteniamo

$$\forall i = 1, \dots, n \quad A\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

Analogamente si dimostra il risultato per $A^\top \mathbf{u}_i = \sigma_i \mathbf{v}_i$ □

Corollario 4.

Hp: Sia $A \in \mathbb{R}^{m \times n}$ con $m \geq n$

Ts: $\|A\|_2 = \sigma_1$ e $\|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$

Dimostrazione. 1)

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|U \Sigma V^\top \mathbf{x}\|_2$$

Dalla **Proposizione ??** sappiamo che le due matrici ortogonali U e V non modificano la norma. Possiamo quindi riscrivere l'uguaglianza come

$$\max_{\|\mathbf{x}\|_2=1} \|\Sigma \mathbf{x}\|_2$$

cioè

$$\max_{\|\mathbf{x}\|_2=1} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

e poiché $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$ otteniamo il massimo per $\mathbf{x} = [1 \ 0 \ \dots \ 0]^\top$ ovvero $\|A\|_2 = \sigma_1$
2)

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |(U\Sigma V^\top)_{ij}|^2}$$

scrivendo il prodotto matriciale elemento per elemento otteniamo che

$$(U\Sigma V^\top)_{ij} = \sum_{k=1}^r U_{ik} \Sigma_{kk} (V^\top)_{kj} = \sum_{k=1}^r U_{ik} \Sigma_{kk} V_{jk}$$

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^r U_{ik} \Sigma_{kk} V_{jk}}$$

Tuttavia per costruzione della matrice Σ

$\Sigma_{kk} = \sigma_k \geq 0 \ \forall k = 1, \dots, r$ con $r \leq n = \min\{m, n\}$ e $\Sigma_{ij} = 0 \ \forall i = 1, \dots, m$ e $\forall j = 1, \dots, n$

Otteniamo la tesi

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

□

Corollario 5. DIMOSTRAZIONE DA CONTROLLARE

Hp: Sia $A, E \in \mathbb{R}^{m \times n}$

Ts: $\sigma_{\max}(A + E) \leq \sigma_{\max}(A) + \|E\|_2$ e $\sigma_{\min}(A + E) \geq \sigma_{\min}(A) - \|E\|_2$

Dimostrazione. Dal Corollario ?? segue che

$$\sigma_{\min}(A) \|\mathbf{x}\|_2 \leq \|A\mathbf{x}\|_2 \leq \sigma_{\max}(A) \|\mathbf{x}\|_2$$

e scegliendo un vettore $\mathbf{x} = 1$ otteniamo che

$$\sigma_{\min}(A) \leq \|A\mathbf{x}\|_2 \leq \sigma_{\max}(A)$$

Applicando questo risultato al nostro caso e la definizione di norma otteniamo

$$\sigma_{\min}(A) - \|E\|_2 \leq \|A\|_2 - \|E\|_2 \leq \|A + E\|_2 \leq \|A\|_2 + \|E\|_2 = \sigma_{\max}(A) + \|E\|_2$$

da cui la tesi. \square

Corollario 6.

Hp: Sia $A \in \mathbb{R}^{m \times n}$ e $\mathbf{z} \in \mathbb{R}^m$

Ts: $\sigma_{\max}([A \mathbf{z}]) \geq \sigma_{\max}(A)$ e $\sigma_{\min}([A \mathbf{z}]) \leq \sigma_{\min}(A)$

Dimostrazione. Sia $A = U\Sigma V^\top$ la decomposizione SVD di A e consideriamo il vettore $\mathbf{x} = \mathbf{v}_1$, ottenuto dalla prima colonna della matrice V , e la matrice $\bar{A} = [A \mathbf{z}] \in \mathbb{R}^{m \times n+1}$. Sfruttando il **Corollario ??** otteniamo

$$\sigma_{\max}(A) = \|A\mathbf{x}\|_2 = \|\bar{A} \begin{bmatrix} x \\ 0 \end{bmatrix}\|_2 \leq \sigma_{\max}(\bar{A})$$

Analogamente si dimostra che $\sigma_{\min}(A) \geq \sigma_{\min}(\bar{A})$ \square

Osservazione: Ricordiamo che possiamo la decomposizione SVD come

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \cdot \mathbf{v}_i^\top$$

1.6.3 Teorema di Eckart-Young

Teorema 4. Teorema di Eckart-Young

Hp: Sia $A \in \mathbb{R}^{m \times n}$ di rango r e sia $k \in \mathbb{N}$ tale che $1 \leq k \leq r$ tale che

$$A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \cdot \mathbf{v}_i^\top$$

è la matrice ottenuta troncando la sommatoria che definisce la decomposizione a valori singolari

Ts: La matrice A_k è soluzione ottima il seguente problema di minimizzazione fra tutte le matrici di rango k

$$\min_{\text{Rg}(B)=k} \|A - B\|_2 = \|A - A_k\|_2$$

Dimostrazione. Notiamo che $A_k = U\Sigma_k V^\top$, possiamo quindi scrivere che

$$A - A_k = U\Sigma V^\top - U\Sigma_k V^\top = U(\Sigma - \Sigma_k)V^\top$$

Invertendo la relazione

$$U^\top (A - A_k) V = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 & \cdots & \sigma_{k+1} & 0 & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sigma_r \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

e per unitarietà della norma delle matrici U e $V \Rightarrow \|A - A_k\|_2 = \sigma_{k+1}$.
Perciò, dobbiamo mostrare che $\forall B_k \in \mathbb{R}^{m \times n}$ tale che $\text{Rg}(B_k) = k$

$$\|A - A_k\|_2 = \sigma_{k+1} \leq \|A - B_k\|_2.$$

Poiché $\text{Rg}(B) = k$

$$\text{Col}(B) = \text{Span}([\mathbf{b}_1 \dots \mathbf{b}_k]) \quad \text{e} \quad \text{Ker}(B) = \text{Row}(B)^\perp = \text{Span}([\mathbf{x}_1 \dots \mathbf{x}_{n-k}]),$$

Per la **Proposizione ??**.4 e per il **Teorema ??**

$$\text{Span}([\mathbf{v}_1 \dots \mathbf{v}_{k+1}]) \cap \text{Span}([\mathbf{x}_1 \dots \mathbf{x}_{n-k}]) \neq 0$$

allora deve esistere una combinazione lineare non banale delle prime $k+1$ colonne di V ,
cioè,

$$\mathbf{w} = \gamma_1 \mathbf{v}_1 + \dots + \gamma_{k+1} \mathbf{v}_{k+1},$$

tale che $Bw = 0$. Senza perdita di generalità, possiamo normalizzare w in modo che $\|w\|_2 = 1$ o (equivalentemente) $\gamma_1^2 + \dots + \gamma_{k+1}^2 = 1$. Pertanto,

$$\|A - B_k\|_2^2 \geq \|(A - B_k)w\|_2^2 = \|Aw\|_2^2 = \gamma_1^2 \sigma_1^2 + \dots + \gamma_{k+1}^2 \sigma_{k+1}^2 \geq \sigma_{k+1}^2.$$

Il risultato segue prendendo la radice quadrata di entrambi i lati della disuguaglianza sopra. \square

Definizione 12. SVD ridotta

Sia A una matrice $m \times n$ con rango r . La decomposizione ai valori singolari (SVD) ridotta di A è data da $A = U' \Sigma' V'^\top$ dove:

$$U' = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{m \times r}, \quad \Sigma' = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \in \mathbb{R}^{r \times r} \quad \text{e} \quad V'^\top = \begin{bmatrix} - & \mathbf{v}_1^\top & - \\ - & \mathbf{v}_2^\top & - \\ & \vdots & \\ - & \mathbf{v}_r^\top & - \end{bmatrix} \in \mathbb{R}^{r \times n}$$

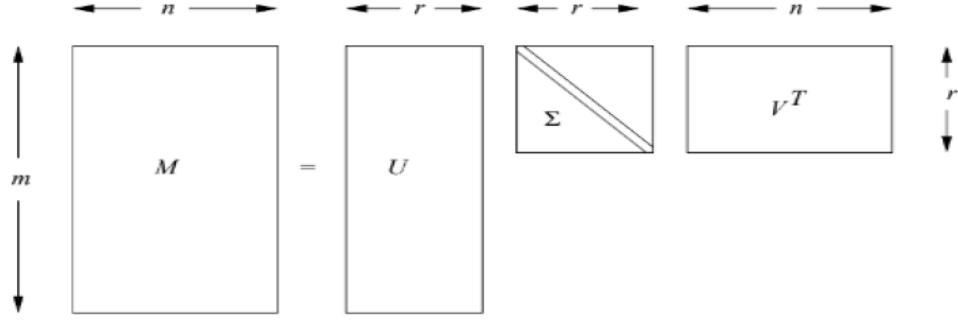


Figura 3: Fattorizzazione SVD Ridotta.

1.7 Applicazioni Teoriche

1.7.1 Pseudoinversa di una Matrice

Definizione 13. Matrice Pseudoinversa o inversa di Moore-Penrose

Sia $A \in \mathbb{R}^{m \times n} \exists A^+ \in \mathbb{R}^{n \times m}$ detta matrice pseudoinversa di A tale che

- $A^+ A \mathbf{x}^+ = \mathbf{x}^+$ per ogni $\mathbf{x}^+ \in \text{Row}(A)$
- $A^+ \mathbf{y}_0 = 0$ per ogni $\mathbf{y}_0 \in \text{Col}(A)^\perp$

Valgono le seguenti uguaglianze:

$$\text{Row}(A^+) = \text{Col}(A) \quad \text{Col}(A^+) = \text{Row}(A) \quad \text{Ker}(A^+) = \text{Col}(A)^\perp$$

La matrice A è la matrice pseudoinversa di A^+ : $A = (A^+)^+$. In particolare

- $AA^+ \mathbf{v} = \mathbf{v}$ per ogni $\mathbf{v} \in \text{Col}(A)$

Osservazione: Se si considera la generica applicazione lineare $\mathcal{L} : \mathbf{W}_1 \rightarrow \mathbf{W}_2$ tra due spazi euclidei di dimensione finita, senza considerare la scelta di una base, si definisce l'applicazione pseudoinversa come l'unica applicazione lineare $\mathcal{L}^+ : \mathbf{W}_2 \rightarrow \mathbf{W}_1$ con le proprietà che $\mathcal{L}^+ \circ \mathcal{L}(x^+) = x^+$ per ogni $x^+ \in \text{Ker}(\mathcal{L})^\perp$ e $\mathcal{L}^+(y_0) = 0$ per ogni $y_0 \in \text{Im}(\mathcal{L})^\perp$. Se A è la matrice che rappresenta \mathcal{L} rispetto a due basi ortonormali di \mathbf{W}_1 e \mathbf{W}_2 , allora A^+ rappresenta \mathcal{L}^+ rispetto a tali basi.

Proposizione 7. Fattorizzazione SVD e matrice pseudoinversa

Hp: Sia $A \in \mathbb{R}^{m \times n}$ e $A = U\Sigma V^T$ la sua decomposizione a valori singolari

Ts: La matrice pseudoinversa di A è $A^+ = V\Sigma^+U^T \in \mathbb{R}^{n \times m}$ dove i valori singolari di A^+ sono i reciproci dei valori singolari di A

Dimostrazione. Vogliamo verificare l'espressione proposta soddisfa tutte le caratteristiche della **Definizione ??**.

Poniamo $C = V\Sigma^+U^T$. Dobbiamo dimostrare che $CA\mathbf{x}^+ = \mathbf{x}^+$ per ogni $\mathbf{x}^+ \in \text{Row}(A)$ e che $C\mathbf{y}_0 = 0$ per ogni $\mathbf{y}_0 \perp \text{Col}(A)$.

Dall'uguaglianza $A^T = V\Sigma^T U^T$ segue che ogni vettore $\mathbf{x}^+ \in \text{Row}(A) = \text{Col}(A^T)$ è della forma $\mathbf{x}^+ = VX^+$ per un unico $X^T \in \text{Row}(\Sigma) = \text{Col}(\Sigma^T)$. Quindi

$$CA = \mathbf{x}^+ = CAVX^+ = V\Sigma^+U^T U\Sigma V^T \mathbf{x}^+ = V\Sigma^+ \Sigma X^+ = VX^+ = \mathbf{x}^+$$

Da $A^T = V\Sigma^T U^T$ segue che ogni vettore $\mathbf{y}_0 \in \text{Ker}(A^T) = \text{Col}(A)^\perp$ è della forma UY_0 per un unico $Y_0 \in \text{Ker}(\Sigma^T)$. Quindi

$$C\mathbf{y}_0 = CUY_0 = V\Sigma^+U^T UY_0 = V\Sigma^+Y_0 = V0 = 0$$

□

Osservazione: Le proprietà della matrice pseudoinversa possono essere equivalentemente riscritte come:

- $AA^+A = A$,
- $A^+AA^+ = A^+$,
- $(AA^+)^T = AA^+$,
- $(A^+A)^T = A^+A$.

da cui è possibile ricavare la seguente formulazione esplicita della matrice pseudoinversa nel caso $\text{Rg}(A)$ massimo: $A^+ = (A^*A)^{-1}A^T$. Per sostituzione facile verificare che $A^+ = V\Sigma^+U^T$

$$\begin{aligned} A^+ &= (A^T A)^{-1} A^T = (V\Sigma U^T U\Sigma V^T)^{-1} V\Sigma U^T \\ &= (V\Sigma^2 V^T)^{-1} V\Sigma U^T = (V^T)^{-1} \Sigma^{-2} V^{-1} V\Sigma U^T \\ &= V\Sigma^{-2} \Sigma U^T = V\Sigma^{-1} U^T \end{aligned}$$

1.7.2 Soluzione ai Minimi Quadrati

Definizione 14. Soluzione ai minimi quadrati

Data una matrice $A \in \mathbb{R}^{m \times n}$ e $\mathbf{b} \in \mathbb{R}^m$ vogliamo trovare $\mathbf{x}^* \in \mathbb{R}^n$ tale che

$$\mathbf{x}^* = \min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2$$

Teorema 5. Soluzione ai Minimi Quadrati di un sistema lineare

Hp: Sia $A \in \mathbb{R}^{m \times n}$ una matrice di rango massimo con $m \geq n$, e sia $\mathbf{b} \in \mathbb{R}^m$.

Ts: Il sistema di equazioni lineari sovradeterminato $A\mathbf{x} = \mathbf{b}$ ha una soluzione ai minimi quadrati \mathbf{x}^* data da

$$\mathbf{x}^* = (A^\top A)^{-1} A^\top \mathbf{b}.$$

Sfruttando la fattorizzazione SVD $A\mathbf{x} = \mathbf{b}$ può essere riscritto come

$$U\Sigma V^\top \mathbf{x} = \mathbf{b}$$

otteniamo

$$\mathbf{x}^* = ((U\Sigma V^\top)^\top U\Sigma V^\top)^{-1} (U\Sigma V^\top)^\top \mathbf{b}$$

$$\mathbf{x}^* = (V\Sigma^\top \Sigma V^\top)^{-1} (V\Sigma^\top U^\top) \mathbf{b}$$

ricordando che date due matrici quadrate Q e M vale $(MQ)^{-1} = Q^{-1}M^{-1}$ e che

$$\Sigma^\top \Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_r^2 \end{bmatrix}$$

è invertibile per definizione di Valori Singolari.

$$\mathbf{x}^* = V(\Sigma^\top \Sigma)^{-1} \Sigma^\top U^\top \mathbf{b}$$

Riscrivo sfruttando la fattorizzazione SVD ridotta

$$\mathbf{x}^* = V'(\Sigma'^\top \Sigma')^{-1} \Sigma'^\top U'^\top \mathbf{b}$$

dove possiamo semplificare

$$(\Sigma^\top \Sigma)^{-1} \Sigma'^\top = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_r^2} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} = \Sigma'^{-1}$$

Il Risultato è quindi

$$\mathbf{x}^* = V' \Sigma'^{-1} U'^\top \mathbf{b}$$

2 Algoritmi per il Calcolo

3 Applicazioni

3.1 PCA - Principal Component Analysis

La Principal Component Analysis (PCA) è una tecnica che trova applicazione nell'ambito della riduzione della dimensionalità e dell'analisi di dati multivariati. L'obiettivo principale di questo metodo è l'identificazione delle "direzioni" lungo cui i dati sono maggiormente distribuiti al fine di determinare efficacemente i parametri delle legge delle variabili aleatorie da cui sono tratte le osservazioni. L'individuazione delle componenti principali è particolarmente utile nel caso di dataset di considerevoli dimensioni, in cui l'applicazione della PCA permette di ridurre la complessità dei dati mantenendo pur mantenendo al contempo la maggior parte delle informazioni rilevanti, consentendo così una più facile visualizzazione e interpretazione.

Ipotizziamo di condurre un certo numero m di esperimenti aleatori le cui realizzazioni possono essere riassunte in un vettore di n elementi. Il dataset prodotto sarà una matrice X di dimensioni (m, n) . Possiamo quindi calcolare il vettore riga medio $\bar{\mathbf{x}}$ sfruttando la media campionaria di un vettore di variabili aleatorie

$$\bar{\mathbf{x}}_j = \frac{1}{m} \sum_{i=1}^m X_{ij} \quad \forall j = 1, \dots, n$$

Una volta ottenuto $\bar{\mathbf{x}}$ costruiamo la matrice media $\bar{X} = [1 \dots 1]^\top \bar{\mathbf{x}}$ di dimensioni (m, n) . Definiamo ora la matrice $B = X - \bar{X}$; questo passaggio equivale ad una traslazione dell'origine del dataset. Infine, sfruttando le proprietà del prodotto matriciale, notiamo che la matrice di covarianza di B è data da

$$C = \frac{1}{n-1} B^\top B$$

dove il fattore $\frac{1}{n-1}$ è dato dalla non distorsione dello stimatore della varianza campionaria.

Per le proprietà matriciali dimostrate precedentemente la matrice C è simmetrica e definita positiva; ciò risulta coerente con quanto ci aspetteremmo da una matrice di covarianza dove $C_{ii} = \lambda_i^2$ è la varianza mentre $C_{ik} = \lambda_{ik}$ la covarianza campionaria fra la i -esima e k -esima colonna.

Date le buone proprietà di C per il **Teorema ??**, possiamo riscrivere l'espressione come segue

$$C = V D V^\top \Rightarrow D = V^\top C V$$

dove gli autovettori sono le componenti principali e gli autovalori le varianze. Intuitivamente, possiamo affermare che la i -esima coppia autovalore-autovettore indica la dispersione del dataset lungo una determinata direzione.

Sappiamo inoltre che la matrice V è la matrice contenente i vettori singolari di destra della matrice B , che possiamo riscrivere come $B = U\Sigma V^\top$. Sostituendo questa relazione otteniamo

$$C = \frac{1}{n-1} B^\top B = \frac{1}{n-1} V \Sigma^\top \Sigma V^\top \Rightarrow D = \frac{1}{n-1} \Sigma^2$$

dove

$$\lambda_k = \frac{\sigma_k^2}{n-1}$$

Illustriamo ora questo concetto dal punto di vista grafico

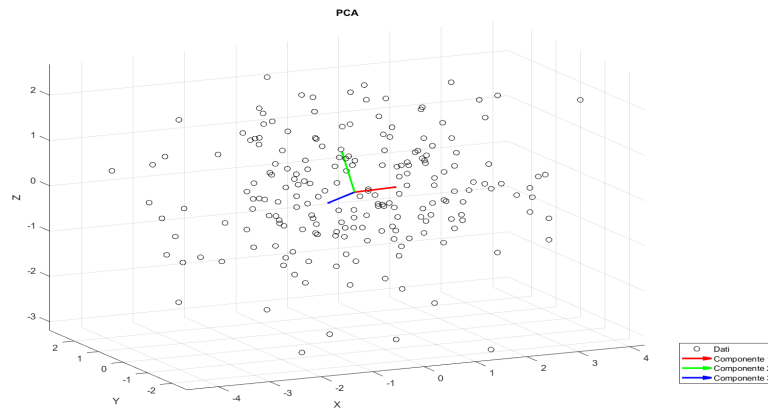


Figura 4: Componenti Principali in \mathbb{R}^3

```
% Parametri della distribuzione gaussiana
mu = [0 0 0]; % Media
sigma = [1 0.5 0.2; 0.5 1 0.3; 0.2 0.3 1]; % Matrice di covarianza
% Generazione del dataset casuale distribuito gaussianamente in R^3
rng(42); % Impostiamo il seed per la replicabilità
n = 400; % Numero di punti nel dataset
data = mvnrnd(mu, sigma, n); % Generiamo i dati casuali
% Calcolo della PCA
coeff = pca(data); % Calcoliamo i coefficienti della PCA
% Plot del dataset e dei vettori della PCA
% Plot dei punti
scatter3(data(:,1), data(:,2), data(:,3), 'MarkerEdgeColor', 'k');
hold on;
% Primo, Secondo e Terzo vettore della PCA
quiver3(mean(data(:,1)), mean(data(:,2)), mean(data(:,3)), ...
```

```

        coeff(1,1), coeff(2,1), coeff(3,1), 'r', 'LineWidth', 2);
quiver3(mean(data(:,1)), mean(data(:,2)), mean(data(:,3)), ...
        coeff(1,2), coeff(2,2), coeff(3,2), 'g', 'LineWidth', 2); A
quiver3(mean(data(:,1)), mean(data(:,2)), mean(data(:,3)), ...
        coeff(1,3), coeff(2,3), coeff(3,3), 'b', 'LineWidth', 2);
% Titoli e grafica
xlabel('X'); ylabel('Y'); zlabel('Z');
title('PCA');
grid on; axis equal;
legend({'Dati', 'Componente 1', 'Componente 2', 'Componente 3'});

```

3.2 Compressione Immagini

La compressione delle immagini tramite la fattorizzazione SVD è un approccio fondamentale nell'ambito della riduzione della dimensionalità dei dati. L'obiettivo principale di questa procedura è quello di ridurre lo spazio di archiviazione richiesto per codificare le informazioni contenute in un'immagine mantenendo al contempo una qualità visiva accettabile.

Dal punto di formale possiamo rappresentare un'immagine in bianco e nero come una matrice X dove a ciascun elemento è associato un pixel dell'immagine con valori da 0 a 1 che rappresentano la scala di grigi che va dal bianco, 0, al nero, 1.

Per il **Teorema ??** siamo in grado di costruire un'approssimazione progressiva della matrice X sfruttando la sommatoria dei valori singolari di destra, sinistra e dei valori singolari minimizzando l'errore commesso.

$$X_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \cdot \mathbf{v}_i^\top$$

Inoltre, sfruttando il **Corollario ??** saremo in grado di stimare l'errore commesso

$$\|X - X_k\|_2 = \left\| \sum_{i=k+1}^r \sigma_i \mathbf{u}_i \cdot \mathbf{v}_i^\top \right\|_2 = \sigma_{k+1}$$

3.3 DMD - Dynamic Mode Decomposition

La Dynamic Mode Decomposition (DMD) è uno strumento per l'analisi dei dati e il rilevamento di strutture coerenti all'interno di essi. La DMD si basa sull'analisi di dataset spazio-temporali composti dall'unione di molteplici osservazioni di un certo numero di grandezze in determinati istanti di tempo. Gli ambiti di applicazione possono variare dalle serie temporali alle immagini dinamiche.

L'approccio si concentra sull'identificazione di strutture coerenti, o modi dominanti,

che emergono nei dati consentendo così la descrizione e la comprensione dei fenomeni sottostanti anche di tipo non lineare e non stazionario. Grazie a queste caratteristiche la DMD risulta essere particolarmente adatta a sistemi complessi.

Dato un insieme di osservazioni di m parametri in n istanti di tempo $\{\mathbf{x}(t_k), \mathbf{x}(t'_k)\}_{k=1}^n$, dove $t'_k = t_k + \Delta t$, costruiamo le matrici X e X' così definite

$$X = \begin{bmatrix} | & | & \dots & | \\ \mathbf{x}(t_1) & \mathbf{x}(t_2) & \dots & \mathbf{x}(t_n) \\ | & | & \dots & | \end{bmatrix} \quad X' = \begin{bmatrix} | & | & \dots & | \\ \mathbf{x}(t'_1) & \mathbf{x}(t'_2) & \dots & \mathbf{x}(t'_n) \\ | & | & \dots & | \end{bmatrix}$$

Assumendo un campionamento uniforme possiamo scrivere che $t_k = k\Delta t$ e di conseguenza quindi vale la seguente relazione $t'_k = t_k + \Delta t = t_{k+1}$. L'obiettivo della DMD si riassume nel trovare la matrice A tale che $X' \simeq AX$ e per l'ipotesi di campionamento uniforme unita alle proprietà del prodotto matriciale segue che $\mathbf{x}_{k+1} \simeq A\mathbf{x}_k$.

Dal punto di vista matematico il problema si traduce nel seguente problema di minimizzazione $A = \min_{A \in \mathbb{R}^{m \times m}} \|X' - AX\|_F = X'X^+$

ricopio dal quad.

Siamo quindi in grado di individuare la soluzione in forma esplicita

$$A = \min_{A \in \mathbb{R}^{m \times m}} \|X' - AX\|_F = X'X^+$$

Una volta ottenuta la matrice quadrata A saremo interessati a trovare i suoi autovalori e autovettori. Tuttavia, nonostante siamo in possesso di una formula esplicita che ci permette calcolarla se consideriamo una matrice $X \in \mathbb{R}^{m \times n}$ dove m sono i parametri osservati e n le realizzazioni temporali, notiamo che la matrice A dovrebbe avere dimensioni $m \times m$. Nelle applicazioni, dove accade di frequente che i parametri osservati sono dell'ordine di 10^6 , ciò si traduce in una matrice con 10^{12} elementi rendendone quindi computazionalmente ingestibile il calcolo.

Per ovviare a questo problema possiamo ricorrere alla POD (proper orthogonal decomposition) in modo da ottenere una matrice \tilde{A} di dimensioni sensibilmente minori, $(n \times n)$, ma che al contempo approssimi al meglio le caratteristiche di $A \in \mathbb{R}^{m \times m}$.

3.3.1 POD - Proper Orthogonal Decomposition

Implementazione dell'Algoritmo

Step 1.

Calcoliamo la fattorizzazione SVD ridotta della matrice $X = U'\Sigma'V'^T$ dove $U' \in \mathbb{R}^{m \times r}$, $\Sigma' \in \mathbb{R}^{r \times r}$ e $V' \in \mathbb{R}^{n \times r}$ con $r \leq n = \min\{m, n\}$. Questa scelta risulta essere cruciale nel caso di elevate dimensioni della matrice X e di $\text{Rg}(X)$ non massimo permettendo di semplificare il costo computazionale dei passaggi che seguiranno.

Ricordiamo che per le proprietà della fattorizzazione SVD ridotta vale la seguente relazione $U'^\top U' = I_r \in \mathbb{R}^{r \times r}$ e $V'^\top V' = I_r \in \mathbb{R}^{r \times r}$ e le colonne della matrice U' prendono il nome di Modi della POD.

Step 2.

Dalla soluzione del problema di minimizzazione sappiamo che la soluzione $A = X'X^+$ può essere scritta come

$$A = X'V'\Sigma'^{-1}U'^\top$$

In questo modo otteniamo una matrice A di dimensioni

$$(m \times m) = (m \times n) \times (n \times r) \times (r \times r) \times (r \times m)$$

Tuttavia come evidenziato in precedenza il nostro interesse si concentra sui primi r autovalori ed autovettori principali di A e per ridurre le dimensioni pur mantenendo la massima "somiglianza" alla matrice di partenza sfruttiamo la POD

$$\bar{A} = U'^\top A U' = U'^\top X'V'\Sigma'^{-1} \in \mathbb{R}^{r \times r}$$

dove è importante notare che gli r autovalori di \bar{A} sono gli stessi r autovalori principali della matrice A . Possiamo quindi trovare direttamente la matrice \bar{A} senza calcolare la matrice estesa A riducendo così sensibilmente il costo computazionale. Una volta introdotta la matrice \bar{A} il modello lineare diventa $\bar{\mathbf{x}}_{k+1} \simeq \bar{A}\bar{\mathbf{x}}_k$ dove $\mathbf{x} = U'\bar{\mathbf{x}}$

Step 3.

Calcoliamo la decomposizione spettrale della matrice \bar{A}

$$\bar{A}W = W\Lambda$$

dove le colonne w_i $i = 1 \dots r$ della matrice $W \in \mathbb{R}^{n \times n}$ sono gli autovettori di \bar{A} mentre $\Lambda \in \mathbb{R}^{n \times n}$ è la matrice diagonale contenente gli autovalori.

Step 4.

I modi $\Phi \in \mathbb{R}^{m \times m}$ sono ricostruiti sfruttando la matrice degli autovettori W del modello lineare di dimensioni ridotte e la matrice dei dati X'

$$\Phi = X'V'\Sigma'^{-1}W$$

notiamo in particolare le dimensioni delle matrici che compongono il prodotto che genera la matrice quadrata finale $(m \times m) = (m \times n) \times (n \times r) \times (r \times r) \times (r \times r)$.

Notiamo che abbiamo quindi ricavato gli autovalori della matrice di partenza A senza calcolarne esplicitamente né il valore né la decomposizione spettrale

$$A\Phi = X'X^+\Phi = X'V'\Sigma'^{-1}U'^\top X'V'\Sigma'^{-1}W$$

e ricordando che la matrice $\bar{A} = U'^\top X'V'\Sigma'^{-1}$ e che $\bar{A}W = W\Lambda$ otteniamo

$$A\Phi = X'V'\Sigma'^{-1}\bar{A}W = X'V'\Sigma'^{-1}W\Lambda = \Phi\Lambda$$

Una volta calcolati i Modi della Dynamic Mode Decomposition sorge spontanea una domanda: data la matrice Φ e Λ siamo in grado di ricostruire [Completo con Ampiezza dei modi b Ricsotruzione dello stato pag 264 e 265](#) fondamentali per l'articolo medico

4 DMD e Modelli Epidemiologici

Come abbiamo osservato nel paragrafo precedente la DMD si presta all'analisi di dati generati da sistemi complessi con un comportamento fortemente non lineare. Un esempio di sistemi di questo tipo sono i modelli pandemici tra cui uno dei più noti è il Modello SIR su cui concentreremo le nostre attenzioni da qui in poi.

5 Bibliografia

Riferimenti bibliografici

- [1] Nome Cognome, Titolo del Libro, Casa Editrice, Anno Di Pubblicazione.
- [2] Nome Cognome, Titolo del Libro, Casa Editrice, Anno Di Pubblicazione.
- [3] Nome Cognome, Titolo del Libro, Casa Editrice, Anno Di Pubblicazione.