5. SVD e Minimi Quadrati Lineari

Analisi Numerica - A.A. 2018/2019

October 19, 2022

sommario

Vettori ortogonali e ortonormali

- I vettori $v_1, v_2, \ldots v_m$ vettori di R^n Si dicono **ortogonali** se:

  $$v_i^T v_j = 0, \quad se\ i \neq j$$

  I vettori $x = (1, 1)$ e $y = (-1, 1) \in R^2$ sono ortogonali.

- I vettori $v_1, v_2, \ldots v_m$ si dicono **ortonormali** se sono ortogonali e di lunghezza unitaria ($v_i^T v_i = \|v_i\| = 1$).

  I vettori $x = \frac{1}{\sqrt{2}}(1, 1)$ e $y = \frac{1}{\sqrt{2}}(-1, 1)$ sono ortonormali.

Matrici ortogonali

Una matrice W di dimensione $n \times n$ si dice **ortogonale** se le sue colonne sono vettori ortonormali.
La matrice

$$W = \begin{pmatrix} 1/\sqrt{3} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1 & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 & 0 \end{pmatrix}$$

Proprieta' di una matrice ortogonale:

1. $W^T = W^{-1}$
2. $\forall x \in R^n, \|x\|_2 = \|Wx\|_2$ quindi W e' una isometria, cioè mantiene le distanze.

Decomposizione in Valori Singolari (SVD)

$m \geq n$

- La Decomposizione in Valori Singolari (SVD) è uno strumento utilissimo, sia dal punto di vista teorico che computazionale.
- Data una matrice A $m \times n$, la SVD costruisce basi ortogonali di R^m e R^n che permettono di rappresentare A attraverso una matrice diagonale.

SVD

Teorema. *Sia A una matrice reale $m \times n$ di rango k, $k \leq n \leq m$. Allora esistono:*

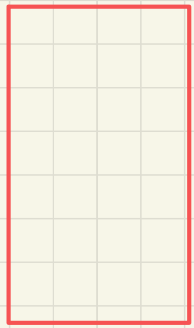
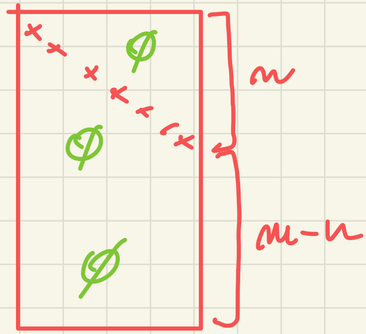
- *una matrice ortogonale U $m \times m$*
- *una matrice ortogonale V $n \times n$*
- *una matrice diagonale Σ $m \times n$*

tali che:

$\sigma_i \to$ valori singolari

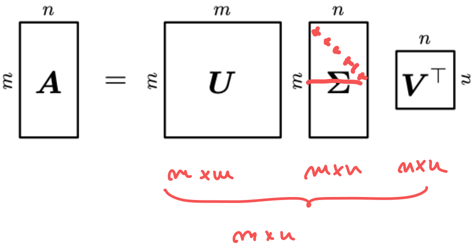
$$A = U\Sigma V^T, \quad \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ \dots & \dots & \dots \\ & & \sigma_n \\ & 0 & \end{pmatrix}$$

$m \geq n$

A_{m}  $=$ 

n

SVD (cont.)



SVD (cont.)

$Ax = \lambda x$

Gli elementi $\sigma_1 \geq \sigma_2 \geq \dots \sigma_n \geq 0$ sono i **valori singolari** di A. Il rango k è uguale al numero di valori singolari positivi, i.e. k=rango(A) sse $\sigma_1 \geq \sigma_2 \dots \geq \sigma_k > 0$ e $\sigma_{k+1} = \dots \sigma_n = 0$.

Le colonne di $U = (\mathbf{u}_1, \mathbf{u}_2, \dots \mathbf{u}_m)$ e di $V = (\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_n)$ sono, rispettivamente, i **vettori singolari sinistri e destri di A** associati ai valori singolari σ_i e formano una base ortonormale di R^m e R^n, rispettivamente, poichè vale la relazione:

$i = 1, \dots m$

$$\boxed{Au_i = \sigma_i v_i}$$

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

$$\Rightarrow U^T A = \underbrace{U^T U}_{I} \Sigma V^T \Rightarrow$$

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

$$A = \underbrace{U \Sigma V^T} = \sum_{i=1}^{n} \sigma_i \underline{u_i} \, \underline{v_i}^T$$

$$U = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad V = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

$$U \cdot V = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} (5, 6) = \begin{pmatrix} 5 & 6 \\ 15 & 18 \end{pmatrix} +$$

$$\begin{pmatrix} 2 \\ 4 \end{pmatrix} (7, 8) = \begin{pmatrix} 14 & 16 \\ 28 & 32 \end{pmatrix}$$

SVD: decomposizione diadica

PCA

Teorema. *Se la matrice reale A $m \times n (m \geq n)$ ha rango k e decomposizione in valori singolari $A = U\Sigma V^T$ (come nel teroema precedente), allora A può essere scritta nella forma diadica come somma di matrici di rango 1 (diadi):*

$$\boxed{A = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T .}$$ → rango = k

$\sigma_{k+1} = \dots \quad \sigma_n = 0$

$\sigma_i u_i v_i^T$ → matrice $m \times n$

DIADE → rango = 1

SVD: risultati

$(AB)^T = B^T A^T$

$V = (V^T)^T \Sigma^T U^T$

Abbiamo la segunete relazione fra i valori singolari di A e gli autovalori di $A^T A$:

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T$$

$I \quad (U^T = U^{-1})$

$$A^\top A = V \Sigma^\top \Sigma V^\top = V \underbrace{\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}}_{\Sigma^T\Sigma} V^\top = V^{-1}.$$

Hence and

$$\boxed{\sigma_i = \sqrt{\lambda_i(A^T A)}, \; i = \ldots, n}$$

λ_i: autovalore di $A^T A$

$$M = X \Delta X^{-1}$$

$X \to$ matrice autovettori X

$$\Delta = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$\to$ autovalori

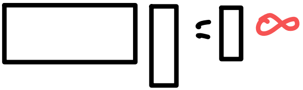
Decomposizione agli autovalori

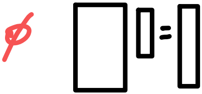
SVD: risultati

In particulare:

- $$\sigma_1 = \sqrt{\lambda_{max}(A^T A)} = \boxed{\sqrt{\rho(A^T A)} = \|A\|_2} = \sigma_1$$
- $$\sigma_k = \sqrt{\lambda_{min}(A^T A)} \Longrightarrow \boxed{\|A^{-1}\| = \frac{1}{\sigma_k}} \quad \frac{1}{\sigma_n}$$
- se m = n
$$\boxed{K_2(A) = \frac{\sigma_1}{\sigma_k}} \qquad K(A) = \|A\| \, \|A^{-1}\|$$
- ~~norma di Frobenius~~:
$$\cancel{\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \dots \sigma_k^2)^{1/2}}$$

Il problema lineare dei minimi quadrati





Vogliamo ora risolvere il sistema

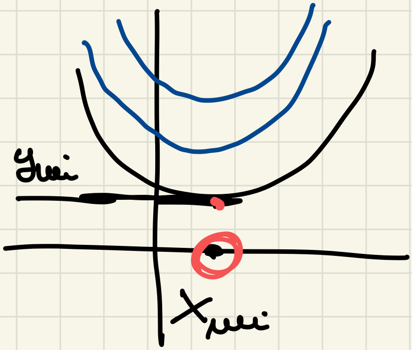
$$Ax = b, \quad A \in R^{m\times n}, \; b \in R^{\overset{m}{M}}, \; x \in R^n \qquad m > n$$

Il sistema in generale **non ha soluzioni**.
Cerchiamo allora di rendere più piccolo possibile il vettore **residuo**:

$$\boxed{r = Ax - b}$$

Il problema lineare dei minimi quadrati (Linear Least Squares LSQ) è formulato come:

$$\boxed{min_x \|\mathbf{b} - A\mathbf{x}\|_2^2 = \min_{x\in R^m} \|r\|_2^2}$$



$$(\arg) \min \|x\| = \min \|x\|^2$$

Il problema lineare dei minimi quadrati

Teorema. *Condizione necessaria e sufficiente affinchè* $\boldsymbol{x}^*$ *sia soluzione del problema LSQ è che soddisfi le equazioni normali:*

$$\boxed{A^T(b - A\mathbf{x}) = 0.} \Rightarrow A^TAx - A^Tb = 0$$

$$\Downarrow$$

$$A^TAx = A^Tb$$

Dimostrazione
Detta $\rho(\mathbf{x}) = \|\mathbf{b} - A\mathbf{x}\|_2^2$, si ha che:

$$\nabla\rho(\mathbf{x}) = -2A^T(\mathbf{b} - A\mathbf{x})$$

Imponendo $\nabla\rho(\mathbf{x}) = 0$ si ha la tesi. $\square$

$$f: R^2 \to R$$

$$\min_{x \in R^2} f \to \nabla f(x) = 0$$

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 0 \\[2ex] \dfrac{\partial f}{\partial x_2} = 0 \end{cases}$$

$$\|r\|_2^2 = r^T \cdot r = 2r^T$$

$$\|Ax - b\|_2^2 = (\overbrace{Ax - b}^{r})^T (Ax - b) =$$

$$2r \frac{dr}{dx} = \cancel{2}A^T(Ax - b) = 0$$

$$A^T(Ax - b) = 0 \quad \text{EQ. NORMALI}$$

Il problema lineare dei minimi quadrati

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

$A^T A$ $n \times n$
simmetrica
semi def. positiva

La matrice delle equazioni normali $A^T A$ (detta anche Gramiano) è quadrata, simmetrica e: $r = \text{rango}(A)$

- se $r = n$ (cioè A ha rango massimo) allora è anche definita positiva e quindi il problema LSQ ha una **UNICA** soluzione.
- se $r < n$ (cioè A NON ha rango massimo) allora $A^T A$ è definita nonnegativa. In questo caso, le soluzioni del problema LSQ generano un sottospazio di R^n di dimensione $n - r$.

Quindi se $r < n$ la soluzione non è unica. Si può però rendere unica imponendo ulteriori condizioni come per esempio che la soluzione abbia minima norma in l2.

Calcolo delle soluzioni di un problema LSQ

Caso di rango massimo =n

$$
\begin{aligned}
||Ax-b||_2^2 &= (Ax-b)^T(Ax-b) \\
&= (-b^T + x^TA^T)(Ax-b) \\
&= -b^TAx + b^Tb + x^TA^TAx - x^TA^Tb \\
&\text{(sommando i termini simili } b^TAx = x^TA^Tb) \\
&= x^TA^TAx - 2x^TAb + b^Tb
\end{aligned}
$$

Si pone:

$$f(x) = x^TA^TAx - 2x^TAb + b^Tb \qquad (2)$$

Calcolo delle soluzioni di un problema LSQ

Per minimizzare la norma bisogna imporre che la derivata prima sia nulla, in questo caso non avendo una sola un'incognita bisogna usare il gradiente, che per una generica funzione è diefinito come segue:

$$f : \mathbb{R}^n \to \mathbb{R} \qquad \nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

Nel caso di (2) si ha:

- $\nabla(x^T A^T A x) = 2A^T A x$
- $\nabla(2x^T A b) = A^T b$
- $\nabla(b^T b) = 0$

Allora:

$$\nabla f(x) = 2A^T A x - A^T b \tag{3}$$

rango $r = n$ (massimo)

quindi la soluzione unica del problema è la soluzione del sistema lineare:

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

Risoluzione di un problema LSQ con SVD

caso $r < n$ (rango non massimo)

Teorema. La soluzione $\hat{\mathbf{x}}^*$ di norma minima del problema LSQ di rango r (che coincide con la soluzione del problema LSQ di rango massimo $r = n$)è data da:

$$\hat{\mathbf{x}}^* = A^+\mathbf{b} = V\Sigma^+U^T\mathbf{b} = \boxed{\sum_{i=1}^{r} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\underbrace{\mathbf{v}_i}_{\in \mathbb{R}^n} = \hat{x}^*} \in \mathbb{R}^n$$

$\mathbb{R}$

e le corrispondenti norme della soluzione e del residuo sono:

$$\|\hat{\mathbf{x}}^*\|_2^2 = \sum_{i=1}^{r}\left(\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\right)^2 \qquad \|\mathbf{r}^*\|_2^2 = \sum_{i=r+1}^{n}(\mathbf{u}_i^T\mathbf{b})^2$$

Risoluzione di un problema LSQ con SVD

FACOLTATIVA

DIMOSTRAZIONE

Infatti moltiplicando per U^T a sinistra:

$$\begin{aligned} ||Ax-b||_2^2 &= ||U^TAx - U^Tb||_2^2 \\ &= ||U^TAVV^Tx - U^Tb||_2^2 \end{aligned}$$

Posto $y = V^Tx \in \mathbb{R}^n$ e $g = U^Tb \in \mathbb{R}^m$
Se $A = U\Sigma V^T \Rightarrow U^TAV = \Sigma$
Allora:

$$\begin{aligned} ||Ax-b||_2^2 &= ||\Sigma y - g||_2^2 \\ &= \sum_{i=1}^{k}(\sigma_i y_i - g_i)^2 + \sum_{i=k+1}^{n} g_i^2 \end{aligned}$$

Per minimizzare tale quantità è sufficiente scegliere:

$$y_i = \frac{g_i}{\sigma_i} = \frac{U^Tb}{\sigma_i} \quad i = 1, \dots, k$$

Risoluzione di un problema LSQ con SVD

FACOLTATIVA

Poichè $x^* = Vy$ risulta:

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

Allora in corrispondenza di tale soluzione, la norma del residuo è:

$$||r||_2^2 = \sum_{i=k+1}^{n} (g_i)^2 = \sum_{i=k+1}^{n} (u_i^T b)^2$$

$\square$ FINE DIMOSTR

Definizione di pseudoinversa

Sia $A \in \mathbb{R}^{m \times n}$, con $rg(A) = k \leq min(m, n)$ e $A = U\Sigma V^T$ la sua decomposizione in valori singolari. Si definisce **pseudoinversa** la matrice:

$$A^+ = V\Sigma^+U^T \qquad dove \quad (\Sigma^+)_{ij} = \begin{cases} \dfrac{1}{\sigma_i} & se\ i = j\ e\ i \leq k \\ 0 & altrimenti \end{cases}$$

Proprietà

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. $(AA^+)^T = AA^+$
4. $(A^+A)^T = A^+A$

La pseudoinversa di una matrice rettangolare A permette di scrivere la soluzione del problema dei minimi quadrati (1) in modo simile alla soluzione $x = A^{-1}b$ di un sistema lineare quadrato, cioè (6) è equivalente a:

$$x^* = V\Sigma^+U^Tb \quad \Longrightarrow \quad \boxed{\mathrm{x}^* = A^+b}$$

Condizionamento di un problema LSQ

Definizione. Il numero di condizione in norma 2 (o numero di condizione spettrale) di una matrice $m \times n$ con valori singolari $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0$ è definito da:

$$cond(A) = \frac{\sigma_1}{\sigma_r}$$

Risultato. Se $\mathbf{x}^*$ è la soluzione del problema LSQ, si ha:

$$\frac{\|\Delta \mathbf{x}\|_2}{\|\mathbf{x}^*\|_2} \leq cond(A)^2 \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2}$$

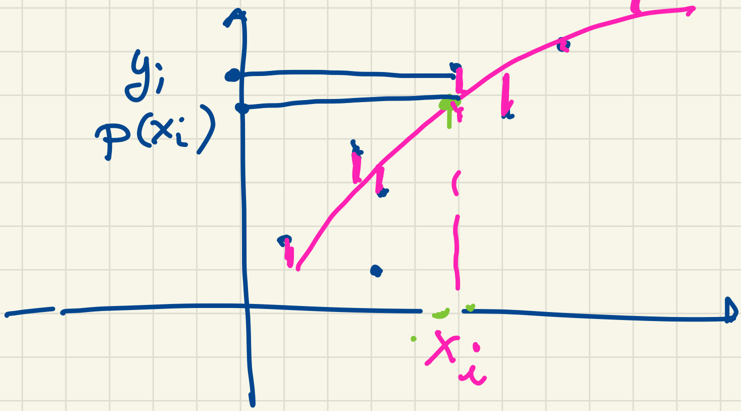
Complessità computazionale e numero di condizione

1. Caso rango massimo:
   - Equazioni normali con Gauss: c.c. $mn^2 + 1/3n^3$, cond: $K(A)^2$
   - Equazioni normali Cholesky ($A = LL^T$): c.c $2mn^2 - 2/3n^3$, cond: $K(L)^2$
2. Caso rango non massimo:
   - SVD: c.c. $4mn^2 + 8n^3$, cond: $K(A)$

REGRESSIONE POLINOMIALE

APPROSSIMARE DATI

$(x_i, y_i) \quad i = 0 \dots m \quad$ ($m+1$ dati)



$$p(x) = c_0 + c_1 x + c_2 x^2 + \dots c_n x^n$$

grado n

$$\underline{c} = \begin{pmatrix} c_0 \\ \vdots \\ c_n \end{pmatrix} \qquad \underline{y} = \begin{pmatrix} y_0 \\ \vdots \\ y_m \end{pmatrix}$$

$\underline{c} \in \mathbb{R}^{n+1} \qquad \underline{y} \in \mathbb{R}^{m+1}$

$$r_i = y_i - p(x_i) \quad \mathbf{|} \quad i = 0 \dots m$$