

0.1 Metodo dei Minimi Quadrati

Supponiamo di avere un sistema di equazioni lineari $A\mathbf{x} = \mathbf{b}$. Sappiamo che un tale sistema è compatibile precisamente quando la colonna dei termini noti appartiene allo spazio delle colonne di A . Se immaginiamo di avere quindi un sistema incompatibile, questo vorrà dire che il vettore differenza $\mathbf{b} - A\mathbf{x}$ è non nullo e possiamo cercare di minimizzare la sua lunghezza. Precisiamo questa idea nella seguente

Definizione 0.1.1. Chiamiamo una *soluzione ai minimi quadrati* del sistema $A\mathbf{x} = \mathbf{b}$ un vettore $\bar{\mathbf{x}} \in \mathbb{R}^n$ tale che

$$\|\mathbf{b} - A\bar{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Al variare di $\mathbf{x}$ in $\mathbb{R}^n$, $A\mathbf{x}$ varia nell'immagine di A , o, detto in altra maniera, $A\mathbf{x}$ varia nello spazio generato dalle colonne di A , che indichiamo con W . Cerchiamo la migliore approssimazione a $\mathbf{b}$ all'interno di W . Segue dal Teorema di approssimazione che la soluzione desiderata è $A\bar{\mathbf{x}} = P_W(\mathbf{b})$. Questo nuovo sistema è sicuramente compatibile perché la colonna dei termini noti è necessariamente nello spazio delle colonne di A . Per risolvere tale sistema lineare, conviene osservare che $\mathbf{b} - A\bar{\mathbf{x}} = \mathbf{b} - P_W(\mathbf{b})$ è ortogonale a W , ossia ortogonale alle colonne di A . In definitiva, quindi, se $\mathbf{a}_i$ è la i -esima colonna di A , abbiamo:

$$\mathbf{a}_i^T (\mathbf{b} - A\bar{\mathbf{x}}) = 0$$

per ogni i . Equivalentemente:

$$A^T (\mathbf{b} - A\bar{\mathbf{x}}) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

ovvero

$$A^T \mathbf{b} - A^T A \bar{\mathbf{x}} = \mathbf{0}$$

ossia

$$A^T A \bar{\mathbf{x}} = A^T \mathbf{b}.$$

Questo nuovo sistema lineare si dice *sistema di equazioni normali* per $\bar{\mathbf{x}}$.

Lemma 0.1.2. Sia $A \in M(m \times n, \mathbb{R})$. Il rango di A è uguale al rango della matrice $A^T A$.

Dimostrazione. $A^T A$ è una matrice quadrata $n \times n$. Il teorema della dimensione ci dice che $n = \text{rg}(A) + \dim(\text{Ker} A)$, ma anche $n = \text{rg}(A^T A) + \dim(\text{Ker}(A^T A))$. Dimostriamo che i ranghi delle due matrici sono uguali provando che i nuclei sono uguali.

Per far ciò, supponiamo che $\mathbf{x} \in \text{Ker} A$. Allora $A\mathbf{x} = \mathbf{0}$ e quindi $A^T A\mathbf{x} = A^T \mathbf{0} = \mathbf{0}$ cioè $\mathbf{x} \in \text{Ker}(A^T A)$. Viceversa: se $A^T A\mathbf{x} = \mathbf{0}$ allora $\mathbf{x}^T A^T A\mathbf{x} = 0$, ma questo è il prodotto scalare di $A\mathbf{x}$ per se stesso e quindi è zero se e solo se $A\mathbf{x} = \mathbf{0}$. $\square$

Corollario 0.1.3. La matrice $A^T A$ è invertibile se e solo se le colonne di A sono linearmente indipendenti.

Abbiamo quindi dimostrato il seguente

Teorema 0.1.4. *Sia $A \in M(m \times n, \mathbb{R})$. Allora:*

- (1) $\bar{\mathbf{x}}$ è una soluzione ai minimi quadrati di $A\mathbf{x} = \mathbf{b}$ se e solo se $\bar{\mathbf{x}}$ è una soluzione del sistema di equazioni normali

$$A^T A \bar{\mathbf{x}} = A^T \mathbf{b}.$$

- (2) Se A ha colonne linearmente indipendenti, la soluzione ai minimi quadrati è unica ed è data da

$$\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

Avendo osservato in precedenza che $A\bar{\mathbf{x}} = P_W(\mathbf{b})$ e che $\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$, nel caso in cui le n colonne di A sono linearmente indipendenti, avremo allora che

$$A(A^T A)^{-1} A^T \mathbf{b} = P_W(\mathbf{b}).$$

In altre parole la matrice

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

è la matrice della trasformazione lineare

$$P_W : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

che proietta ortogonalmente $\mathbb{R}^m$ sul sottospazio W delle colonne di A .

Esercizio 0.1.5. Calcolare la matrice della proiezione ortogonale di $\mathbb{R}^3$ sul sottospazio W costituito dal piano di equazione cartesiana $x - y + 2z = 0$.

Soluzione Una base di W è costituita da $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$. Consideriamo la matrice rettangolare che ha questi vettori per colonne:

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Allora

$$A^T A = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Quindi

$$(A^T A)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$$

e infine

$$A(A^T A)^{-1} A^T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{5}{6} & \frac{1}{6} & -\frac{1}{3} \\ \frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

è la matrice desiderata.

Esempio 0.1.6. Si può verificare che, posto

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 3 \\ 1 & -1 & 2 \\ 1 & 0 & 1 \end{pmatrix},$$

la matrice $A(A^T A)^{-1} A^T$ è

$$\frac{1}{4} \begin{pmatrix} 3 & 1 & 1 & -1 \\ 1 & 3 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{pmatrix}$$

Definizione 0.1.7. La matrice $A^+ = (A^T A)^{-1} A^T$ si dice *pseudoinversa* di A .

Esempio 0.1.8. Calcolare la soluzione approssimata ai minimi quadrati del sistema incompatibile $A\mathbf{x} = \mathbf{b}$ dove

$$A = \begin{pmatrix} 1 & 5 \\ 2 & -2 \\ -1 & 1 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}.$$

Calcoliamo

$$A^T A = \begin{pmatrix} 1 & 2 & -1 \\ 5 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 2 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 30 \end{pmatrix}$$

e

$$A^T \mathbf{b} = \begin{pmatrix} 1 & 2 & -1 \\ 5 & -2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 16 \end{pmatrix}$$

Le equazioni normali sono allora il sistema

$$\begin{pmatrix} 6 & 0 \\ 0 & 30 \end{pmatrix} \bar{x} = \begin{pmatrix} 2 \\ 16 \end{pmatrix}$$

da cui ricaviamo $\bar{x} = \begin{pmatrix} \frac{1}{3} \\ \frac{16}{30} \end{pmatrix}$.

Il metodo dei minimi quadrati fu ideato da Gauss per calcolare l'orbita dell'asteroide Cerere scoperto nel 1801. Da allora si è rivelato uno strumento utilissimo per trattare dati sperimentali.

Supponiamo, a titolo di esempio, di avere n punti $(x_1, y_1), \dots, (x_n, y_n)$, ottenuti da un ipotetico esperimento e di aver motivo di credere che i valori di x e y siano legati da una relazione lineare del tipo $y = a + bx$. Una tale funzione, se esistesse, soddisferebbe:

$$a + bx_1 = y_1$$

$$\begin{aligned} a + bx_2 &= y_2 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

ovvero $A\mathbf{x} = \mathbf{b}$, avendo posto

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Tale sistema, quasi sicuramente non ammette alcuna soluzione $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, tuttavia la soluzione ai minimi quadrati fornisce la cosiddetta *retta ai minimi quadrati*, che ha il pregio di minimizzare la norma del vettore *errore* $\mathbf{e} = \mathbf{b} - A\mathbf{x}$.

Esercizio 0.1.9. Determinare la retta ai minimi quadrati, relativa ai tre punti $(1, 2)$, $(2, 2)$, $(3, 4)$.

Soluzione Dobbiamo calcolare la soluzione ai minimi quadrati del sistema:

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}.$$

Le relative equazioni normali sono:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$$

cioè

$$\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 8 \\ 18 \end{pmatrix}$$

da cui $a = \frac{2}{3}$, $b = 1$, da cui $y = \frac{2}{3} + x$ è la desiderata retta ai minimi quadrati.

Nell'esercizio che segue cerchiamo una parabola che meglio approssima quattro punti dati. Il problema, come vedremo, è del tutto analogo al precedente.

Esercizio 0.1.10. Siano dati $P_1 = (-2, 4)$, $P_2 = (-1, 1)$, $P_3 = (1, 1)$, $P_4 = (2, -6)$. Cerchiamo una curva di equazione

$$y = a_0 + a_1x + a_2x^2$$

tale che

$$\begin{cases} a_0 - 2a_1 + 4a_2 = 4 \\ a_0 - a_1 + a_2 = 1 \\ a_0 + a_1 + a_2 = 1 \\ a_0 + 2a_1 + 4a_2 = -6 \end{cases}$$

Il sistema ha matrice

$$\begin{pmatrix} 1 & -2 & 4 & 4 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & -6 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 4 & 4 \\ 0 & 1 & -3 & -3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -10 \end{pmatrix}$$

da cui esso risulta incompatibile.

Consideriamo allora

$$\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b},$$

con

$$A = \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}, \quad A^T A = \begin{pmatrix} 2 & 0 & 5 \\ 0 & 5 & 0 \\ 5 & 0 & 17 \end{pmatrix}, \quad (A^T A)^{-1} = \frac{1}{90} \begin{pmatrix} 85 & 0 & -25 \\ 0 & 9 & 0 \\ -25 & 0 & 10 \end{pmatrix}.$$

Si ha quindi:

$$\bar{\mathbf{x}} = \frac{1}{90} \begin{pmatrix} 85 & 0 & -25 \\ 0 & 9 & 0 \\ -25 & 0 & 10 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -2 & -1 & 1 & 2 \\ 4 & 1 & 1 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \\ 1 \\ -6 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 5 \\ -6 \\ -2 \end{pmatrix},$$

dunque $a_0 = \frac{5}{3}$, $a_1 = -2$, $a_2 = -\frac{2}{3}$ e

$$y = \frac{5}{3} - 2x - \frac{2}{3}x^2$$

è la richiesta parabola ai minimi quadrati.

Esempio 0.1.11. L'andamento della popolazione di Roma dopo l'unità d'Italia è data dalla seguente Tabella¹

Anno	pop. (in migliaia)
1871	209,222
1881	269,813
1901	416,028
1911	511,076
1921	650,258
1931	916,858
1951	1626,793
1961	2155,093
1971	2739,952
1981	2797,337
1991	2733,908
2001	2546,804
2011	2617,175

¹<http://www.tuttitalia.it/lazio/33-roma/statistiche/censimenti-popolazione/>

Vogliamo trovare la legge lineare che meglio approssima questo andamento.
Le matrici A e $\mathbf{b}$ sono

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 10 \\ 1 & 30 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 80 \\ 1 & 90 \\ 1 & 100 \\ 1 & 110 \\ 1 & 120 \\ 1 & 130 \\ 1 & 140 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 209,222 \\ 269,813 \\ 416,028 \\ 511,076 \\ 650,258 \\ 916,858 \\ 1626,793 \\ 2155,093 \\ 2739,952 \\ 2797,337 \\ 2733,908 \\ 2546,804 \\ 2617,175 \end{pmatrix}$$

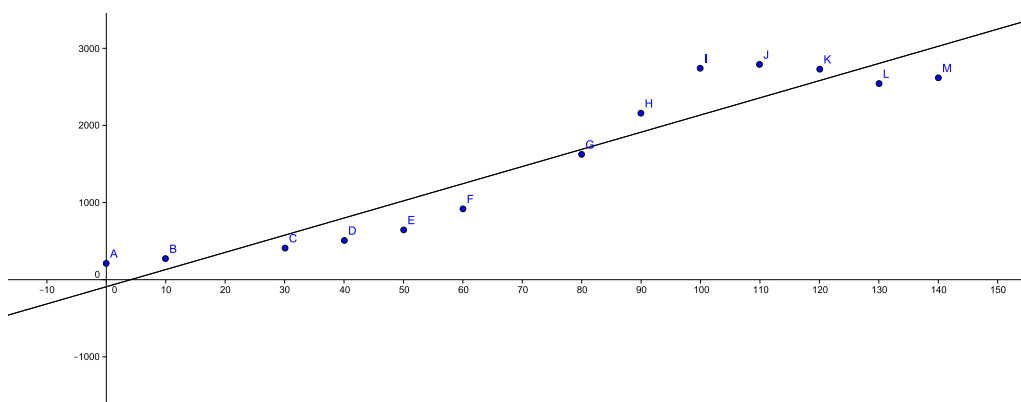
Allora $A^T A = \begin{pmatrix} 13 & 960 \\ 960 & 96200 \end{pmatrix}$. L'inversa di questa matrice è

$$(A^T A)^{-1} = \begin{pmatrix} \frac{481}{1645} & -\frac{24}{8225} \\ \frac{24}{8225} & -\frac{13}{329000} \end{pmatrix}$$

La soluzione si trova infine calcolando

$$(A^T A)^{-1} A^T \mathbf{b} = \begin{pmatrix} -91,245 \\ 22,2672 \end{pmatrix}$$

In altre parole, la retta desiderata è $y = 22,2672x - 91,245$, (v. figura)²



²I grafici relativi a questo esercizio sono stati ottenuti, come molti altri in questo libro, con GeoGebra: www.geogebra.org. In particolare qui si è usato il comando RegLin o Regressione

Secondo questo modello di crescita, nel 2021 la popolazione di Roma dovrebbe essere all'incirca, sostituendo $x = 150$ nell'equazione della retta, 3.248.835 persone.

Se si sceglie un modello con un polinomio di quarto grado, invece di un modello lineare, si ottiene la funzione polinomiale $f(x) = 252.3313768224035 - 3.502186383790344x + 0.088876374230785x^2 + 0.004991144677039x^3 - 0.000033170972662x^4$ che per $x = 150$ fornisce il valore 1.779.030: drammaticamente diverso. Quale modello sarà più fedele alla realtà? Ovviamente la risposta esula dal corso di Geometria. (v. figura)

