Minimi quadrati

Formulazione e risoluzione del problema

In queste poche righe si vuole dare un'idea della risoluzione dei problemi ai minimi quadrati, si cercherà di trattare la problematica in modo semplice e senza l'utilizzo di pesanti nozioni teoriche, arrivando a formulare la risoluzione ripercorrendo le linee guida utilizzate dai grandi matematici che nel passato formularono per primi il problema.

La formulazione di un problema ai minimi quadrati può essere vista in modo molto semplice utilizzando esempi concreti. L'unica abilità richiesta è la capacità di utilizzare il semplice calcolo simbolico alle sommatorie e l'utilizzo delle matrici ed il relativo calcolo.

Prima di divertirci ☺, immaginiamo di avere a disposizione una serie di misure che le possiamo rappresentare come un vettore $\bar{y} = [y_1 \quad y_2 \quad \cdots \quad y_m]^T$ (dove la ^{T} sta a significare TRASPOSTA, in quanto si immagina che il vettore sia un vettore colonna). Come dato assodato, ciascuna misura è soggetta a rumore, ossia: vi è presenza di un elemento di disturbo che va a "falsare" il valore letto della misura. Immaginiamo anche di conoscere a priori il modello che è alla base del sistema che abbiamo misurato. Facciamo un'ulteriore ipotesi, imponiamo che il nostro sistema sia lineare!

Federico Milan
Cittadella (PD)
Italy
federico@plcforum.it

Da questa introduzione, sperando di non aver annoiato nessuno, possiamo estrapolare un modello simbolico del nostro problema, che non è altro che un sistema lineare di m equazioni.

$$\begin{cases} a_{11}\cdot x_1 + a_{12}\cdot x_2 + \ldots + a_{1n}\cdot x_n = y_1 \\ a_{21}\cdot x_1 + a_{22}\cdot x_2 + \ldots + a_{2n}\cdot x_n = y_2 \\ \vdots \\ a_{m1}\cdot x_1 + a_{m2}\cdot x_2 + \ldots + a_{mn}\cdot x_n = y_m \end{cases}$$

Il sistema indica che l'ipotetica misura y_1 è funzione lineare di n parametri e di n variabili indipendenti x. L'ultimo sforzo che possiamo fare è di trasformare il sistema lineare in forma matriciale come qui sotto.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \text{ o equivalentemente } A\cdot\overline{x} = \overline{y}$$

Arrivati a modellizzare il problema con questa scrittura si aprono delle magiche porte, ossia, qualunque sia il significato delle y misure possiamo costruire una comune teoria di risoluzione. La potenza del motodo è proprio nell'automatizzare il processo di calcolo ☺.

Data la veste matriciale, il problema si traduce nel trovare un vettore x, tale che riesca a minimizzare l'errore, più precisamente, vogliamo che l'errore quadratico sia minimizzato. Questo si traduce nel trovare un $\hat{x}$ tale che $\left\|A\cdot\hat{x}-\overline{y}\right\|^2$ sia tendente a zero, o almeno il più piccolo possibile.

Creandoci una notazione possiamo dire che vogliamo che $\varepsilon^2 = \left\|A\cdot\hat{x}-\overline{y}\right\|^2$ sia il più piccolo possibile.

Dobbiamo a questo punto fare una precisazione, cosa significa la simbologia $\|.\|$? Si aprono vari campi, ma a noi interessa applicare una definizione semplice e intuitiva. Per noi assume significato di modulo e in particolare, ricordando che stiamo parlando di calcoli vettoriali (x,y sono vettori e A è una matrice), il modulo lo potremmo implementare come una distanza Euclidea. Associare $\|.\|$ a una distanza euclidea significa ottenere la seguente eguaglianza: $\|h-r\| = \sqrt{(h_1-r_1)^2 + (h_2-r_2)^2 + \ldots + (h_t-r_t)^2}$. Per inciso, è matematica elementare in quanto se abbiamo il caso particolare t=2, otteniamo la regola di Pitagora ☺!

Ora, stabilito questo abbiamo il mezzo per costruire la funzione di errore ε^2. Abbiamo detto che vogliamo minimizzare l'errore, questo significa che vogliamo trovare quando la

Federico Milan
Cittadella (PD)
Italy

federico@plcforum.it

nostra funzione $\varepsilon^2 = \|A \cdot \hat{x} - \overline{y}\|^2$ assume valore minimo. Possiamo supporre che, se il vettore $\hat{x} \to \pm\infty$ anche $A \cdot \hat{x} - \overline{y} \to \pm\infty$, quindi possiamo ipotizzare che esista un punto di minimo nella funzione ε^2. Per trovare il punto di minimo basta trovare dove si annulla la derivata di $\frac{\partial \varepsilon^2}{\partial \hat{x}} = 0$.

Arrivati a questo punto abbiamo solo un problema, ossia come rappresentare la funzione di errore, in modo da poterne calcolare la derivata? Dobbiamo ricordare cosa è il prodotto tra una matrice e un vettore!
Il prodotto tra una matrice ed un vettore è fatto sommando il prodotto di ciascun elemento di una riga della matrice per ciascun elemento del vettore colonna (Vedi il sistema prima descritto e confrontarlo con il sistema matriciale ☺).

Detto questo possiamo costruire la seguente relazione:

$$\varepsilon^2 = \|A \cdot \hat{x} - \overline{y}\|^2 = \sum_{t=1}^{m} \left(\sum_{z=1}^{n} a_{tz} \cdot \hat{x}_z - y_t \right)^2 \;!$$

Con questa formulazione possiamo calcolare la derivate della funzione di errore. Ricordando la regole di derivazione (che del resto la derivata è un operatore lineare, ossia la derivata di una somma è la somma delle derivate ☺) otteniamo che:

$$\frac{\partial \varepsilon^2}{\partial \hat{x}} = \sum_{t=1}^{m} 2 \cdot a_{tz} \cdot \left(\sum_{z=1}^{n} a_{tz} \cdot \hat{x}_z - y_t \right)$$

Se il calcolo della derivata fa paura basta semplificare è procedere così:

$$\sum_{t=1}^{2} \left(\sum_{z=1}^{2} a_{tz} \cdot \hat{x}_z - y_t \right)^2 = \left(\sum_{z=1}^{2} a_{1z} \cdot \hat{x}_z - y_1 \right)^2 + \left(\sum_{z=1}^{2} a_{2z} \cdot \hat{x}_z - y_2 \right)^2 =$$

$$(a_{11} \cdot \hat{x}_1 + a_{12} \cdot \hat{x}_2 - y_1)^2 + (a_{21} \cdot \hat{x}_1 + a_{22} \cdot \hat{x}_2 - y_2)^2$$

$$= \frac{\partial((a_{11} \cdot \hat{x}_1 + a_{12} \cdot \hat{x}_2 - y_1)^2 + (a_{21} \cdot \hat{x}_1 + a_{22} \cdot \hat{x}_2 - y_2)^2)}{\partial \hat{x}_1} =$$

$$= 2 \cdot a_{11} \cdot (a_{11} \cdot \hat{x}_1 + a_{12} \cdot \hat{x}_2 - y_1) + 2 \cdot a_{11} \cdot (a_{21} \cdot \hat{x}_1 + a_{22} \cdot \hat{x}_2 - y_2)$$

Che porta alla formulazione semplificata del calcolo simbolico.

Trovata l'espressione analitica della derivata della funzione errore, dobbiamo uguagliarla a zero: $\frac{\partial \varepsilon^2}{\partial \hat{x}} = \sum_{t=1}^{m} 2 \cdot a_{tz} \cdot \left(\sum_{z=1}^{n} a_{tz} \cdot \hat{x}_z - y_t \right) = 0$. Per risolvere, possiamo percorrere una strada inversa, ossia: $\sum_{t=1}^{m} 2 \cdot a_{tz} \cdot \left(\sum_{z=1}^{n} a_{tz} \cdot \hat{x}_z - y_t \right) = \sum_{t=1}^{m} 2 \cdot a_{tz} \cdot (A \cdot \hat{x} - y) = 0$. Ora la sommatoria rimanente si comporta come il prodotto di due matrici, ma con la prima matrice trasposta (ossia invertita righe con colonne), e possiamo scrivere:

$$\sum_{t=1}^{m} 2 \cdot a_{tz} \cdot \left(\sum_{z=1}^{n} a_{tz} \cdot \hat{x}_z - y_t \right) = \sum_{t=1}^{m} 2 \cdot a_{tz} \cdot (A \cdot \hat{x} - y) = 2 \cdot A^T \cdot (A \cdot \hat{x} - y) = 0 .$$

Siamo ormai arrivata allo souluzione del problema ☺!
Come procedere? Semplice:

$$2 \cdot A^T \cdot (A \cdot \hat{x} - y) = 2 \cdot \left(A^T \cdot A \cdot \hat{x} - A^T y\right) = 0 \rightarrow A^T \cdot A \cdot \hat{x} = A^T y$$

Ora abbiamo risolto il problema, se non si riesce a vedere la soluzione, basta fare questo ultimo passaggio:

$$A^T \cdot A \cdot \hat{x} = A^T y \rightarrow \hat{x} = \left(A^T \cdot A\right)^{-1} \cdot A^T y$$

Ora siamo arrivati alla definizione del problema ☺!

Come utilizzare questo risultato? Una piccola ricerca su PLC Forum e si scopre su come utilizzarlo ☺!

Per chi odia le ricerche, o per chi si voglia rilassarsi un po' e vedere come sfruttare quanto trovato – per fortuna ci sono i Computer ai giorni d'oggi –possiamo costruirci il nostro problema.

Immaginiamo di avere una serie di misure, ottenute su un modello.

Immaginiamo di conoscere il nostro modello (es.: distanza percorsa da una macchina a velocità costante)

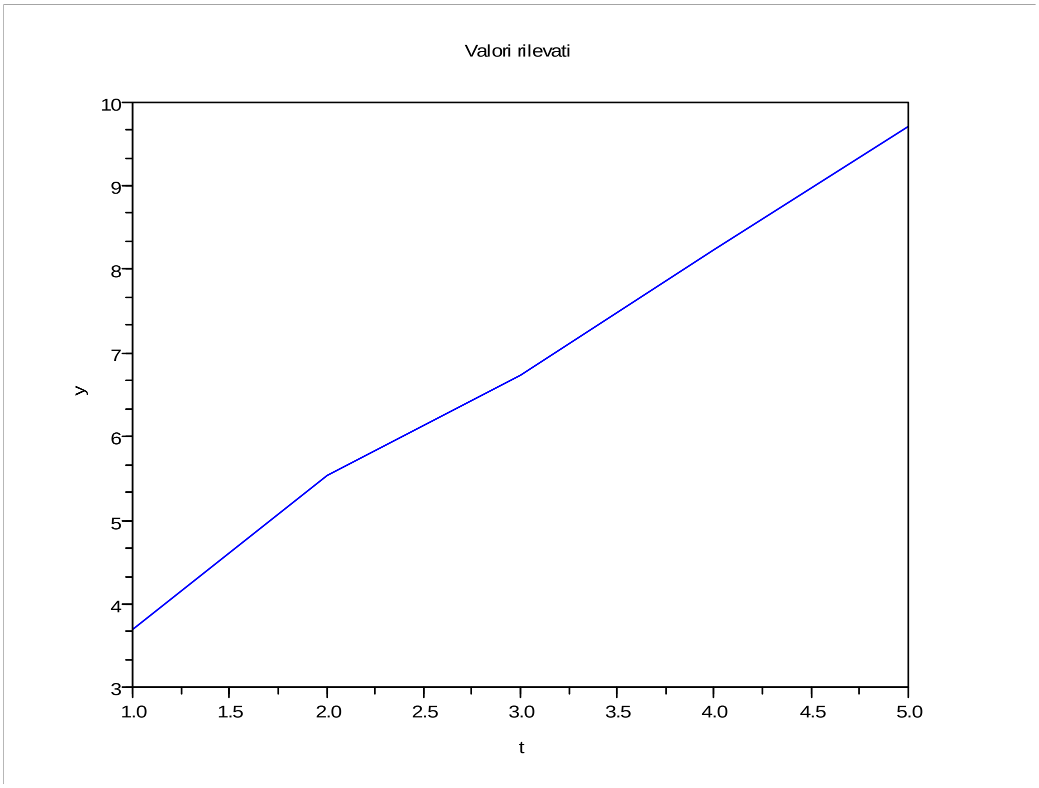
$y = x_1 \cdot t + x_2 + rumore(t)$ *, ma di non conoscere i parametri* x_1 *e* x_2*.*

La misura è soggetta a rumore, ossia ad un disturbo (per es. rumore sull'amplificatore, o a errori di trascrizione dell'operatore di turno!).

Immaginiamo ora di aver rilevato le seguenti misure:

$$y = [3.6985144 \quad 5.5442573 \quad 6.7320748 \quad 8.2312237 \quad 9.7164633]^T$$

che può essere rappresentata dal seguente grafico:



Sappiamo che il modello è lineare. Vogliamo trovare x (il vettore x contiene i parametri cercati)

Possiamo costruirci qualcosa del genere:

$$y = A \cdot x \rightarrow \begin{bmatrix} 3.6985144 \\ 5.5442573 \\ 6.7320748 \\ 8.2312237 \\ 9.7164633 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix} \cdot x$$

Ora applicando le regole prima viste possiamo scrivere:

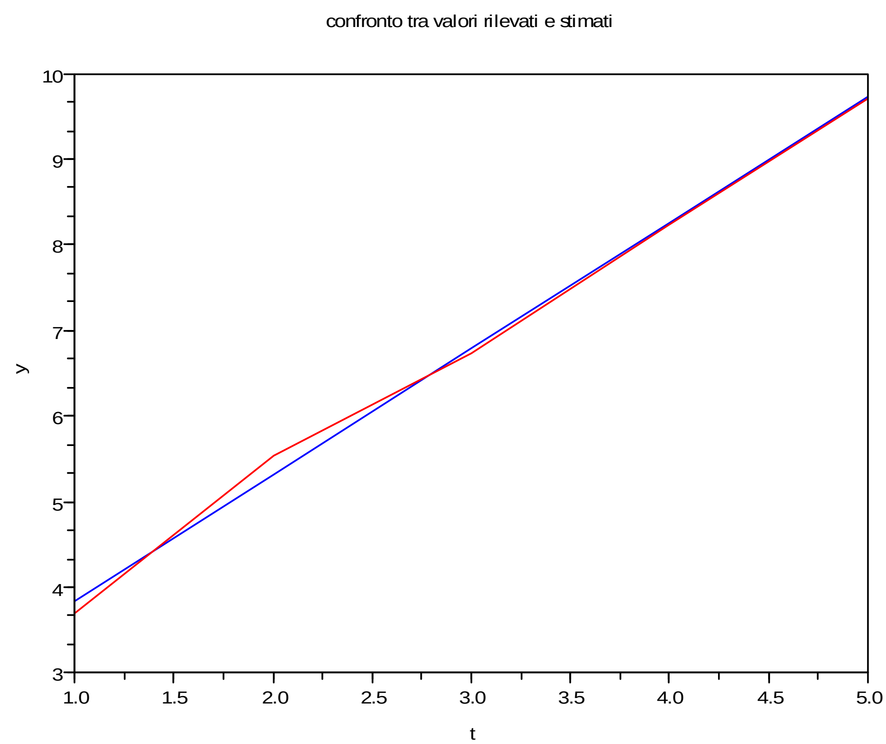
$$A^T \cdot A = \begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix}$$

$$\left(A^T \cdot A\right)^{-1} = \begin{bmatrix} 0.1 & -0.3 \\ -0.3 & 1.1 \end{bmatrix}$$

$$\left(A^T \cdot A\right)^{-1} \cdot A^T = \begin{bmatrix} -0.2 & -0.1 & 0 & 0.1 & 0.2 \\ 0.8 & 0.5 & 0.2 & -0.1 & -..4 \end{bmatrix}$$

$$x = \begin{bmatrix} -0.2 & -0.1 & 0 & 0.1 & 0.2 \\ 0.8 & 0.5 & 0.2 & -0.1 & -..4 \end{bmatrix} \cdot \begin{bmatrix} 3.6985144 \\ 5.5442573 \\ 6.7320748 \\ 8.2312237 \\ 9.7164633 \end{bmatrix} = \begin{bmatrix} 1.4722864 \\ 2.3676474 \end{bmatrix}$$

Ottenendo la seguente rappresentazione grafica:



Siamo riusciti a stimare i parametri che meglio approssimano il nostro modello, modello ricavato da misure sogette a rumore. I parametri reali erano 1.5 e 2. Diciamo che abbiamo avuto poche misure a disposizione, e che il rumore aggiunto non è proprio trascurabile ☺. Ad ogni modo, dato le misure, i parametri torvati sono quelli che approssimano in modo migliore le misure effettuate!

Federico Milan
Cittadella (PD)
Italy
federico@plcforum.it