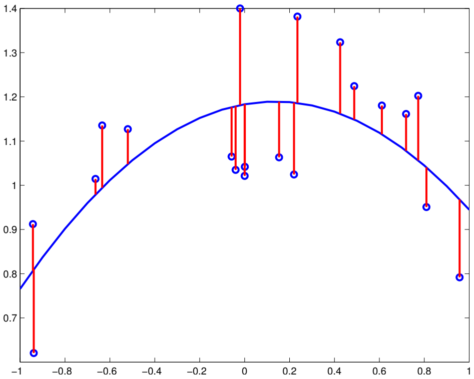
Approssimazione secondo il criterio dei minimi quadrati (caso discreto)

Dati $m+1$ punti distinti (punti di osservazione) $x_0, x_2, \ldots, x_m \in [a, b]$ e $m+1$ valori (osservazioni) $y_0, y_2, \ldots, y_m \in \mathbb{R}$, si vuole determinare il modello matematico che meglio approssima tale insieme di dati sperimentali

$$f(x;\ \alpha_0, \alpha_1, \ldots, \alpha_n) = y$$

dove $\alpha_0, \alpha_1, \ldots, \alpha_n$ sono gli $n+1$ parametri da determinare, con $m > n$.
Si vogliono determinare i parametri incogniti $\alpha_0, \alpha_1, \ldots, \alpha_n$ (detti gradi di libertà) in modo tale che la distanza tra $f(x;\ \alpha_0, \alpha_1, \ldots, \alpha_n)$ e i dati osservati sia minima, ossia più piccola possibile. Occorre poi verificare la bontà del modello adottato.



Approssimazione secondo il criterio dei minimi quadrati (caso discreto)

Se si considera il sistema di $m+1$ equazioni nelle $n+1$ incognite $\alpha_0, \alpha_1, \ldots, \alpha_n$, si ottiene il **sistema sovradeterminato** $m > n$

$$\begin{cases} f(x_1;\ \alpha_0, \alpha_1, \ldots, \alpha_n) - y_1 = \varepsilon_0 \\ f(x_2;\ \alpha_0, \alpha_1, \ldots, \alpha_n) - y_2 = \varepsilon_1 \\ \vdots \\ f(x_m;\ \alpha_0, \alpha_1, \ldots, \alpha_n) - y_m = \varepsilon_m \end{cases}$$

dove ε_i, $i = 0, \ldots, m$, è il disturbo (rumore) che “sporca” l’i-esimo dato y_i.

Principio dei minimi quadrati

Si tratta di determinare $\alpha_0, \alpha_1, \dots, \alpha_n$ in modo tale che ε_i, $i = 0, \dots, m$, siano più piccoli possibile.
Sia $m \geqslant n$: se si considera il principio dei minimi quadrati, si determinano α_0, α_1, $\dots$, α_n in modo che il vettore $\boldsymbol{\varepsilon} = (\varepsilon_0, \dots, \varepsilon_m)^T \in \mathbb{R}^{m+1}$ abbia norma euclidea minima (cioè più piccola possibile), ossia in modo che sia minima la funzione $Q(\alpha_0, \alpha_1, \dots, \alpha_n)$ detta myAlertsomma dei quadrati dei residui:

$$Q(\alpha_0, \alpha_1, \dots, \alpha_n) = \sum_{i=0}^{m} \varepsilon_i^2 = \sum_{i=0}^{m} \Big(f(x_i;\ \alpha_0, \alpha_1, \dots, \alpha_n) - y_i \Big)^2$$

oppure

$$= \sum_{i=0}^{m} w_i \varepsilon_i^2 = \sum_{i=0}^{m} w_i \Big(f(x_i;\ \alpha_0, \alpha_1, \dots, \alpha_n) - y_i \Big)^2$$

dove $w_i > 0$ sono numeri scelti come "pesi" delle osservazioni.
Una possibile scelta dei pesi è $w_i = 1/e_i^2$, dove e_i è una stima approssimata dell'errore sul dato i-esimo.

Si fa l'ipotesi che x_i non siano affetti da errore. In questo caso i moduli dei residui, $|\varepsilon_i|$, misurano la "distanza verticale" tra $\big(x_i, y_i\big)$ e $\big(x_i, f(x_i;\ \alpha_0, \alpha_1, \dots, \alpha_n)\big)$.

Approssimazione lineare

Se $f(x)$ è combinazione lineare di $n+1$ funzioni elementari $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ (che si assume siano continue e derivabili, con derivata continua in $[a, b]$, ossia siano sufficientemente regolari), allora

$$f(x) = \alpha_0\varphi_0(x) + \dots + \alpha_n\varphi_n(x) = \sum_{j=0}^{n} \alpha_j \varphi_j(x) \qquad \Big(\text{per es. } \varphi_j(x) = x^j\Big)$$

e il problema di approssimazione è lineare. Si tratta di trovare $\alpha_0, \dots, \alpha_n$ tali che la funzione $Q(\alpha_0, \dots, \alpha_n)$ sia minima:

$$Q(\alpha_1, \dots, \alpha_n) = \sum_{i=0}^{m} w_i \left(\sum_{j=0}^{n} \alpha_j \varphi_j(x_i) - y_i \right)^2$$

Se si indica con $\boldsymbol{y} = (y_0, \dots, y_m)^T \in \mathbb{R}^{m+1}$ il vettore dei dati, con $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_n)^T \in \mathbb{R}^{n+1}$ il vettore dei parametri da determinare, con $A = (a_{ij})$ la matrice di $m+1$ righe e $n+1$ colonne, detta matrice di regressione lineare, così definita

$$A = \begin{pmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \varphi_n(x_1) \\ \vdots & \vdots & \vdots \\ \varphi_0(x_m) & \varphi_1(x_m) & \varphi_n(x_m) \end{pmatrix}$$

Approssimazione lineare

e con $D = \text{diag}(\sqrt{w_i})_{i=0,\dots,m}$, allora il vettore degli errori è definito come

$$\varepsilon = D(A\boldsymbol{\alpha} - \boldsymbol{y})$$

e dunque la funzione $Q(\alpha_0, \dots, \alpha_n)$, che rappresenta la sua norma euclidea al quadrato, si può scrivere anche come

$$\begin{aligned} Q(\alpha_0, \dots, \alpha_n) &= \|D(A\boldsymbol{\alpha} - \boldsymbol{y})\|_2^2 \\ &= \Big(D(A\boldsymbol{\alpha} - \boldsymbol{y})\Big)^T \Big(D(A\boldsymbol{\alpha} - \boldsymbol{y})\Big) \\ &= \boldsymbol{\alpha}^T A^T D^2 A\boldsymbol{\alpha} - 2\boldsymbol{y}^T D^2 A\boldsymbol{\alpha} + \boldsymbol{y}^T D^2 \boldsymbol{y} \end{aligned}$$

La funzione $\boldsymbol{\alpha}^T A^T D^2 A\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^T D^2 \boldsymbol{y} + \boldsymbol{y}^T D^2 \boldsymbol{y}$ è una forma quadratica associata alla matrice $A^T D^2 A$.

Esempio

Determinare i parametri α_0 e α_1 del modello $f(x) = \alpha_0 \log(x) + \alpha_1$, considerando come dati $(0.5, 5)$, $(1, 5)$, $(4, 1)$ e $(7, 0.1)$.
Si assume che i primi due dati abbiano un errore pari a 0.1 mentre gli ultimi due abbiano errore superiore e pari a 0.5.

In questo caso $m + 1 = 4$ ed $n + 1 = 2$ (dunque $m = 3$ ed $n = 1$). Infatti occorre calcolare i due parametri α_0 e α_1 in modo che siano minimi possibile gli errori dati dalle 4 espressioni:

$$\begin{aligned} \alpha_0 \log(0.5) + \alpha_1 - 5 &= \varepsilon_0 \\ \alpha_0 \log(1) + \alpha_1 - 5 &= \varepsilon_1 \\ \alpha_0 \log(4) + \alpha_1 - 1 &= \varepsilon_2 \\ \alpha_0 \log(7) + \alpha_1 - 0.1 &= \varepsilon_3 \end{aligned}$$

Esempio

Usando il criterio dei minimi quadrati, occorre trovare α_0 e α_1 in modo che la **norma euclidea pesata** del vettore degli errori sia minima possibile:

$$Q(\alpha_0,\alpha_1) = \frac{1}{0.1^2}\left(\alpha_0\log(0.5)+\alpha_1-5\right)^2 + \frac{1}{0.1^2}\left(\alpha_0\log(1)+\alpha_1-5\right)^2 + \frac{1}{0.5^2}\left(\alpha_0\log(4)+\alpha_1-1\right)^2 + \frac{1}{0.5^2}\left(\alpha_0\log(7)+\alpha_1-0.1\right)^2$$

Si osserva che la funzione si può esprimere anche nel seguente modo:

$$Q(\alpha_0,\alpha_1) = \left\| \begin{matrix} \frac{1}{0.1}\left(\alpha_0\log(0.5)+\alpha_1-5\right) \\ \frac{1}{0.1}\left(\alpha_0\log(1)+\alpha_1-5\right) \\ \frac{1}{0.5}\left(\alpha_0\log(4)+\alpha_1-1\right) \\ \frac{1}{0.5}\left(\alpha_0\log(7)+\alpha_1-0.1\right) \end{matrix} \right\|^2 = \left\| \begin{pmatrix} \frac{1}{0.1} & 0 & 0 & 0 \\ 0 & \frac{1}{0.1} & 0 & 0 \\ 0 & 0 & \frac{1}{0.5} & 0 \\ 0 & 0 & 0 & \frac{1}{0.5} \end{pmatrix} \left(\begin{pmatrix} \log(0.5) & 1 \\ \log(1) & 1 \\ \log(4) & 1 \\ \log(7) & 1 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} - \begin{pmatrix} 5 \\ 5 \\ 1 \\ 0.1 \end{pmatrix} \right) \right\|^2$$

Esempio

Ponendo

$$D = \begin{pmatrix} \frac{1}{0.1} & 0 & 0 & 0 \\ 0 & \frac{1}{0.1} & 0 & 0 \\ 0 & 0 & \frac{1}{0.5} & 0 \\ 0 & 0 & 0 & \frac{1}{0.5} \end{pmatrix};\ A = \begin{pmatrix} \log(0.5) & 1 \\ \log(1) & 1 \\ \log(4) & 1 \\ \log(7) & 1 \end{pmatrix};\ \boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix};\ \boldsymbol{y} = \begin{pmatrix} 5 \\ 5 \\ 1 \\ 0.1 \end{pmatrix}$$

la funzione da minimizzare rispetto ai parametri si scrive come

$$Q(\alpha_0,\alpha_1) = \left\|D(A\boldsymbol{\alpha}-\boldsymbol{y})\right\|^2 = \left(D(A\boldsymbol{\alpha}-\boldsymbol{y})\right)^T\left(D(A\boldsymbol{\alpha}-\boldsymbol{y})\right) = (A\boldsymbol{\alpha}-\boldsymbol{y})^T D^2 (A\boldsymbol{\alpha}-\boldsymbol{y})$$

Si può anche porre $\tilde{A} = DA = \begin{pmatrix} \frac{\log(0.5)}{0.1} & \frac{1}{0.1} \\ \frac{\log(1)}{0.1} & \frac{1}{0.1} \\ \frac{\log(4)}{0.5} & \frac{1}{0.5} \\ \frac{\log(7)}{0.5} & \frac{1}{0.5} \end{pmatrix}$ e $\tilde{\boldsymbol{y}} = D\boldsymbol{y} = \begin{pmatrix} \frac{5}{0.1} \\ \frac{5}{0.1} \\ \frac{1}{0.5} \\ \frac{0.1}{0.5} \end{pmatrix}$, da cui

$$Q(\alpha_0,\alpha_1) = \left\|\tilde{A}\boldsymbol{\alpha}-\tilde{\boldsymbol{y}}\right\|^2 = \left(\tilde{A}\boldsymbol{\alpha}-\tilde{\boldsymbol{y}}\right)^T\left(\tilde{A}\boldsymbol{\alpha}-\tilde{\boldsymbol{y}}\right)$$

Formulazione del problema

Il problema si riformula come il problema di trovare la soluzione di

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{n+1}} \|D A\boldsymbol{\alpha} - D\boldsymbol{y}\|_2^2 = \min_{\boldsymbol{\alpha}\in\mathbb{R}^{n+1}} \|\widetilde{A}\boldsymbol{\alpha} - \widetilde{\boldsymbol{y}}\|^2$$

dove $\widetilde{A} = DA$ e $\widetilde{\boldsymbol{y}} = D\boldsymbol{y}$.

Condizione necessaria perchè $\boldsymbol{\alpha}^* = \left(\alpha_0^*, \ldots, \alpha_n^*\right)^T$ renda minima $Q(\alpha_0, \ldots, \alpha_n)$ è che il gradiente di $Q(\alpha_0, \ldots, \alpha_n)$ sia nullo per $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$.

Formulazione del problema

Si deriva allora la funzione $\sum_{i=0}^{m} w_i \left(\sum_{j=0}^{n} \varphi_j(x_i)\alpha_j - y_i\right)^2$ rispetto alle variabili α_j, $j = 0, \ldots, m$, e si pone ciascuna derivata uguale a zero:

$$\frac{\partial}{\partial \alpha_k}\left(\sum_{i=0}^{m} w_i \left(\sum_{j=0}^{n} \varphi_j(x_i)\alpha_j - y_i\right)^2\right) = 0$$

$$2\sum_{i=0}^{m} w_i \left(\sum_{j=0}^{n} \varphi_j(x_i)\alpha_j - y_i\right)\varphi_k(x_i) = 0$$

$$\sum_{i=0}^{m} w_i \sum_{j=0}^{n} \varphi_j(x_i)\varphi_k(x_i)\alpha_j - \sum_{j=0}^{m} w_i y_i \varphi_k(x_i) = 0$$

$$\sum_{j=0}^{n} \alpha_j \sum_{i=0}^{m} w_i\varphi_j(x_i)\varphi_k(x_i) = \sum_{i=0}^{m} w_i y_i \varphi_k(x_i) \quad k = 0, \ldots, n$$

$$\begin{cases} \alpha_0 \sum_{i=0}^{m} w_i\varphi_0(x_i)\varphi_0(x_i) + \alpha_1 \sum_{i=0}^{m} w_i\varphi_0(x_i)\varphi_1(x_i) + \ldots + \alpha_n \sum_{i=0}^{m} w_i\varphi_0(x_i)\varphi_n(x_i) = \sum_{i=0}^{m} w_i y_i \varphi_0(x_i) \\ \alpha_0 \sum_{i=0}^{m} w_i\varphi_1(x_i)\varphi_0(x_i) + \alpha_1 \sum_{i=0}^{m} w_i\varphi_1(x_i)\varphi_1(x_i) + \ldots + \alpha_n \sum_{i=0}^{m} w_i\varphi_1(x_i)\varphi_n(x_i) = \sum_{i=0}^{m} w_i y_i \varphi_1(x_i) \\ \vdots \\ \alpha_0 \sum_{i=0}^{m} w_i\varphi_n(x_i)\varphi_0(x_i) + \alpha_1 \sum_{i=0}^{m} w_i\varphi_n(x_i)\varphi_1(x_i) + \ldots + \alpha_n \sum_{i=0}^{m} w_i\varphi_n(x_i)\varphi_n(x_i) = \sum_{i=0}^{m} w_i y_i \varphi_n(x_i) \end{cases}$$

Formulazione del problema

dove la matrice del sistema $B \in \mathbb{R}^{(n+1)\times(n+1)}$ ha componenti $b_{jk} = \sum_{i=0}^m w_i\varphi_j(x_i)\varphi_k(x_i)$ e il termine noto ha componenti $d_j = \sum_{i=0}^m w_i y_i \varphi_j(x_i)$.

$$B = \begin{pmatrix} \sum_{i=0}^m w_i\varphi_0(x_i)\varphi_0(x_i) & \sum_{i=0}^m w_i\varphi_0(x_i)\varphi_1(x_i) & \dots & \sum_{i=0}^m w_i\varphi_0(x_i)\varphi_n(x_i) \\ \sum_{i=0}^m w_i\varphi_1(x_i)\varphi_0(x_i) & \sum_{i=0}^m w_i\varphi_1(x_i)\varphi_1(x_i) & \dots & \sum_{i=0}^m w_i\varphi_1(x_i)\varphi_n(x_i) \\ \vdots & \vdots & & \vdots \\ \sum_{i=0}^m w_i\varphi_n(x_i)\varphi_0(x_i) & \sum_{i=0}^m w_i\varphi_n(x_i)\varphi_1(x_i) & \dots & \sum_{i=0}^m w_i\varphi_n(x_i)\varphi_n(x_i) \end{pmatrix} \quad \boldsymbol{d} = \begin{pmatrix} \sum_{i=0}^m w_i y_i\varphi_0(x_i) \\ \sum_{i=0}^m w_i y_i\varphi_1(x_i) \\ \vdots \\ \sum_{i=0}^m w_i y_i\varphi_n(x_i) \end{pmatrix}$$

Si può osservare che $B = A^T D^2 A$ e $\boldsymbol{d} = A^T D^2 \boldsymbol{y}$.

$$A = \begin{pmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \dots & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \dots & \varphi_n(x_1) \\ \vdots & \vdots & & \vdots \\ \varphi_0(x_m) & \varphi_1(x_m) & \dots & \varphi_n(x_m) \end{pmatrix}$$

Pertanto il sistema delle equazioni normali è

$$A^T D^2 A\boldsymbol{\alpha} = A^T D^2 \boldsymbol{y} \qquad \widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} = \widetilde{A}^T\widetilde{\boldsymbol{y}}$$

o, se $D = I_{m+1}$,

$$A^T A\boldsymbol{\alpha} = A^T\boldsymbol{y}.$$

Formulazione del problema

Il sistema di equazioni normali è un sistema di $n+1$ equazioni in $n+1$ incognite, con B semidefinita positiva (forma quadratica semidefinita positiva).
Pertanto se $\boldsymbol{\alpha}^*$ è punto di minimo, ossia soluzione di

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{n+1}} \|\widetilde{A}\boldsymbol{\alpha} - \widetilde{\boldsymbol{y}}\|_2^2 = \min_{\boldsymbol{\alpha}\in\mathbb{R}^{n+1}} Q(\alpha_0, \dots, \alpha_n) \qquad (1)$$

allora $\boldsymbol{\alpha}^*$ è soluzione di

$$\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} = \widetilde{A}^T\widetilde{\boldsymbol{y}} \qquad (2)$$

Viceversa si può mostrare che se $\boldsymbol{\alpha}^*$ è soluzione del sistema di equazioni normali, allora è punto di minimo, ossia è soluzione di (1).

Formulazione del problema

Infatti, se $\boldsymbol{\alpha}^*$ è soluzione di (2), $\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha}^* = \widetilde{A}^T\widetilde{\boldsymbol{y}}$, oppure $\widetilde{\boldsymbol{y}}^T\widetilde{A} = \boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}$, allora vale che

$$\begin{aligned}
Q(\boldsymbol{\alpha}) - Q(\boldsymbol{\alpha}^*) &= \boldsymbol{\alpha}^T\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} - 2\widetilde{\boldsymbol{y}}^T\widetilde{A}\boldsymbol{\alpha} + \widetilde{\boldsymbol{y}}^T\widetilde{\boldsymbol{y}} - \boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha}^* + 2\widetilde{\boldsymbol{y}}^T\widetilde{A}\boldsymbol{\alpha}^* \\
&\quad - \widetilde{\boldsymbol{y}}^T\widetilde{\boldsymbol{y}} \\
&= \boldsymbol{\alpha}^T\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} - \boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha}^* + 2\widetilde{\boldsymbol{y}}^T\widetilde{A}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) \\
&= \boldsymbol{\alpha}^T\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} - \boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha}^* + 2\boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) \\
&= \boldsymbol{\alpha}^T\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} + \boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha}^* - 2\boldsymbol{\alpha}^{*T}\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} \\
&= (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T\widetilde{A}^T\widetilde{A}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \geqslant 0
\end{aligned}$$

Per cui $Q(\boldsymbol{\alpha}) \geqslant Q(\boldsymbol{\alpha}^*)$. Dunque $\boldsymbol{\alpha}^*$ è punto di minimo. Se $\widetilde{A}^T\widetilde{A}$ è definita positiva, allora $\boldsymbol{\alpha}^*$ è punto di minimo proprio ($Q(\boldsymbol{\alpha}) > Q(\boldsymbol{\alpha}^*)$ per $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$).
Pertanto i due problemi (1) e (2) sono equivalenti.

In sintesi, abbiamo dimostrato che la soluzione del problema di miglior approssimazione lineare secondo il criterio dei minimi quadrati si trova studiando le soluzioni del sistema

$$\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} = \widetilde{A}^T\widetilde{\boldsymbol{y}}$$

Formulazione del problema

Vale che

$$0 = \widetilde{A}^T\widetilde{\boldsymbol{y}} - \widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} = \widetilde{A}^T(\widetilde{\boldsymbol{y}} - \widetilde{A}\boldsymbol{\alpha}) = \widetilde{A}^T\boldsymbol{r}$$

Nell'esempio, si tratta di risolvere il sistema

$$\widetilde{A}^T\widetilde{A}\boldsymbol{\alpha} = \widetilde{A}^T\widetilde{\boldsymbol{y}}$$

Da cui si ottiene $\alpha_0 = -1.2525$ e $\alpha_1 = 4.4917$.

Interpretazione geometrica

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{n+1}} \|A\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2 \qquad A \in \mathbb{R}^{(m+1)\times(n+1)}, \quad \boldsymbol{y} \in \mathbb{R}^{m+1}$$

Sia $S(A)$ il sottospazio di $\mathbb{R}^{m+1}$ generato dalle colonne della matrice A.

- Se $\boldsymbol{y} \in S(A)$, allora si tratta di trovare le coordinate di $\boldsymbol{y}$ rispetto all'insieme di generatori rappresentato dalle colonne della matrice A (il problema diventa di interpolazione, poiché $\min \|A\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2 = 0$).
- Se $\boldsymbol{y} \notin S(A)$, si tratta di determinare $\widehat{\boldsymbol{y}} \in S(A)$ tale che $\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|_2^2$ è minimo. Tale $\widehat{\boldsymbol{y}}$ è la proiezione ortogonale di $\boldsymbol{y}$ su $S(A)$ ed è univocamente determinato. $\boldsymbol{r} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$ è un vettore ortogonale a $S(A)$, ossia $\boldsymbol{r} \perp$ a tutte le colonne di A, $A^T\boldsymbol{r} = \boldsymbol{0}$. Ogni altro $\boldsymbol{y}^* \in S(A)$ è tale che

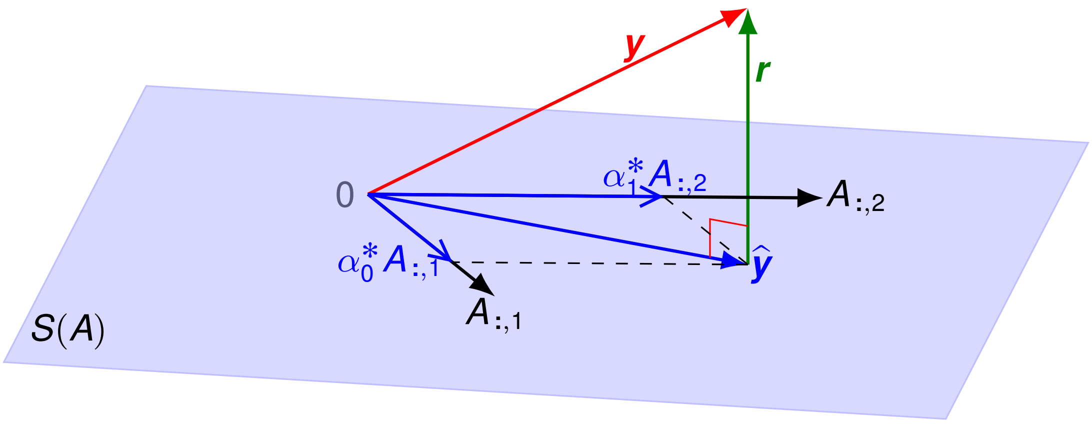
$$\begin{aligned}\|\boldsymbol{y} - \boldsymbol{y}^*\|^2 &= \|\boldsymbol{y} - \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{y}} - \boldsymbol{y}^*\|_2^2 = (\boldsymbol{y} - \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{y}} - \boldsymbol{y}^*)^T(\boldsymbol{y} - \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{y}} - \boldsymbol{y}^*)\\ &= (\boldsymbol{y} - \widehat{\boldsymbol{y}})^T(\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\widehat{\boldsymbol{y}} - \boldsymbol{y}^*)(\widehat{\boldsymbol{y}} - \boldsymbol{y}^*) + 2\,\underbrace{(\widehat{\boldsymbol{y}} - \boldsymbol{y}^*)^T}_{\in S(A),\text{se }\widehat{\boldsymbol{y}}\neq\boldsymbol{y}^*}\,\underbrace{(\boldsymbol{y} - \widehat{\boldsymbol{y}})}_{=\boldsymbol{r}}\\ &= \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|_2^2 + \|\widehat{\boldsymbol{y}} - \boldsymbol{y}^*\|_2^2 > \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|_2^2\end{aligned}$$

Interpretazione geometrica

Poiché $A^T\boldsymbol{r} = \boldsymbol{0}$, $A^T(\boldsymbol{y} - \widehat{\boldsymbol{y}}) = \boldsymbol{0}$, $A^T\boldsymbol{y} = A^T\widehat{\boldsymbol{y}}$

$$\widehat{\boldsymbol{y}} \in S(A) \Longrightarrow \exists \boldsymbol{\alpha}^* \text{ tale che } \widehat{\boldsymbol{y}} = A\boldsymbol{\alpha}^* \Longrightarrow A^T\boldsymbol{y} = A^TA\boldsymbol{\alpha}^*$$

e $\boldsymbol{\alpha}^*$ è la soluzione del problema.
Se le colonne di A sono linearmente indipendenti allora $\boldsymbol{\alpha}^*$ è unica.
Se le colonne di A sono linearmente dipendenti e $\mathrm{rank}(A) = k$, esistono ∞^{n+1-k} vettori $\boldsymbol{\alpha}^*$ tali che $A\boldsymbol{\alpha}^* = \widehat{\boldsymbol{y}}$. Tra queste infinite soluzioni si considera quella di norma euclidea minima.



Osservazione

L'approssimazione lineare secondo il criterio dei minimi quadrati ha una e una sola soluzione $\iff$ A è di rango pieno $\iff$ i vettori $\begin{pmatrix} \varphi_j(x_0) \\ \vdots \\ \varphi_j(x_m) \end{pmatrix}$ sono linearmente indipendenti. In tal caso la soluzione vale

$$\boldsymbol{\alpha}^* = (A^T A)^{-1} A^T y = A^+ \boldsymbol{y}$$

Questo accade se $\varphi_0, \ldots, \varphi_n$ sono funzioni linearmente indipendenti sui punti distinti $x_0, \ldots, x_m$.
L'approssimazione lineare secondo il criterio dei minimi quadrati allora ha una sola soluzione comunque siano scelti i punti $x_0, \ldots, x_m$ purchè distinti in $[a, b]$ se le funzioni $\varphi_0, \ldots, \varphi_n$ formano una matrice a colonne linearmente indipendenti.

Approssimazione polinomiale

Se $\varphi_j(x) = x^j$, ossia nel caso di approssimazione polinomiale, $x^0 = 1, x, \ldots, x^n$ sono sempre linearmente indipendenti su $m + 1 \geqslant n + 1$ punti comunque scelti in $[a, b]$ purché **distinti**.
L'approssimazione lineare polinomiale ha sempre una e una sola soluzione.
Se $m = n$, la soluzione è il polinomio di interpolazione di grado n relativo a $x_0, \ldots, x_n$.
Se $m > n$, la soluzione è il polinomio di grado n con coefficienti che sono soluzione di

$$A^T D^2 A \boldsymbol{\alpha} = A^T D^2 \boldsymbol{y} \quad \text{o, equivalentemente,} \quad B\boldsymbol{\alpha} = \boldsymbol{d}$$

dove

$$A^T D^2 A = \begin{pmatrix} \sum_{i=0}^m w_i & \sum_{i=0}^m w_i x_i & \ldots & \sum_{i=0}^m w_i x_i^n \\ \sum_{i=0}^m w_i x_i & \sum_{i=0}^m w_i x_i^2 & \ldots & \sum_{i=1}^m w_i x_i^{n+1} \\ \vdots & \vdots & & \vdots \\ \sum_{i=0}^m w_i x_i^n & \sum_{i=0}^m w_i x_i^{n+1} & \ldots & \sum_{i=0}^m w_i x_i^{2n} \end{pmatrix}, \quad A^T D^2 \boldsymbol{y} = \begin{pmatrix} \sum_{i=0}^m w_i y_i \\ \sum_{i=0}^m w_i y_i x_i \\ \vdots \\ \sum_{i=0}^m w_i y_i x_i^n \end{pmatrix}$$

o, equivalentemente,

$$b_{jk} = \sum_{i=0}^{m} w_i x_i^{j+k-2}, \qquad d_j = \sum_{i=0}^{m} w_i y_i x_i^{j-1} \qquad j,k = 1,\dots,n+1.$$

N.B. Si ricorda che la numerazione di righe e colonne di matrici e vettori inizia da 1, mentre le numerazioni dei punti x_i e delle funzioni $\varphi_j(x)$ iniziano entrambe da zero.

La matrice di regressione lineare risulta infatti

$$A = \begin{pmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \dots & x_m^n \end{pmatrix} \in \mathbb{R}^{(m+1)\times(n+1)}$$

Esempi

- $n = 0$. $f(x) = \alpha_0$.

$$A = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \Rightarrow \quad A^T D^2 A = \sum_{i=0}^{m} w_i\,, \quad A^T D^2 \mathbf{y} = \sum_{i=0}^{m} w_i y_i \quad \Rightarrow \quad \alpha_0 = \frac{\sum_{i=0}^{m} w_i y_i}{\sum_{i=0}^{m} w_i}$$

  α_0 è il baricentro dei dati.

- $n = 1$. $f(x) = \alpha_0 + \alpha_1 x$. Questo è il caso che permette di costruire la retta di regressione:

$$A = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \Rightarrow \quad A^T D^2 A = \begin{pmatrix} \sum_{i=0}^{m} w_i & \sum_{i=0}^{m} w_i x_i \\ \sum_{i=0}^{m} w_i x_i & \sum_{i=0}^{m} w_i x_i^2 \end{pmatrix}, \quad A^T D^2 \mathbf{y} = \begin{pmatrix} \sum_{i=0}^{m} w_i y_i \\ \sum_{i=0}^{m} w_i y_i x_i \end{pmatrix}$$

  Se $D^2 = I_{m+1}$, allora

$$A^T A = \begin{pmatrix} m+1 & \sum_{i=0}^{m} x_i \\ \sum_{i=0}^{m} x_i & \sum_{i=0}^{m} x_i^2 \end{pmatrix}, \qquad A^T \mathbf{y} = \begin{pmatrix} \sum_{i=0}^{m} y_i \\ \sum_{i=0}^{m} y_i x_i \end{pmatrix}$$

$$\alpha_0 = \frac{\left(\sum x_i^2\right)\left(\sum y_i\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{(m+1)\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}, \qquad \alpha_1 = \frac{(m+1)\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{(m+1)\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}.$$

Poichè A^TA è simmetrica definita positiva, un modo naturale per risolvere il sistema delle equazioni normali è usare la fattorizzazione di Cholesky:

- calcolo di $B = A^TA$ e $\boldsymbol{d} = A^T\boldsymbol{y}$;
- fattorizzazione di $B = LL^T$;
- risoluzione dei sistemi $L\boldsymbol{z} = \boldsymbol{d}$ e $L^T\boldsymbol{\alpha} = \boldsymbol{z}$.

Il costo è di $mn^2/2 + n^3/6 + n^2$ prodotti e altrettante somme. Dal teorema di condizionamento dei sistemi lineari, per la soluzione calcolata $\widetilde{\alpha}$ si ha:

$$\frac{\|\Delta\widetilde{\boldsymbol{\alpha}}\|_2}{\|\widetilde{\boldsymbol{\alpha}}\|_2} \leqslant \frac{\mu_2(B)}{1 - \mu_2(B)\dfrac{\|\Delta B\|_2}{\|B\|_2}} \left(\frac{\|\Delta B\|_2}{\|B\|_2} + \frac{\|\Delta \boldsymbol{d}\|_2}{\|\boldsymbol{d}\|_2}\right)$$

Di più, rispetto alle perturbazioni ΔA sui valori iniziali di A, assunto che $\eta = \|\Delta A\|_2/\|A\|_2$ sia piccolo, si ha:

$$\frac{\|\Delta\widetilde{\boldsymbol{\alpha}}\|}{\|\widetilde{\boldsymbol{\alpha}}\|} \leqslant \eta\Big(\mu_2(A) + \mu_2^2(A)\tan(\theta)\Big) + \mathcal{O}(\eta^2) \quad \text{e} \quad \frac{\|\Delta\boldsymbol{r}\|}{\|\boldsymbol{y}\|} \leqslant 2\eta\mu_2(A) + \mathcal{O}(\eta^2)$$

dove θ è tale che $\sin(\theta) = \|\boldsymbol{r}\|_2/\|\boldsymbol{y}\|_2$, essendo $\boldsymbol{r} = \boldsymbol{y} - A\widetilde{\alpha}$ il residuo. Infine, vale che

$$\mu_2(A^TA) = \frac{\lambda_{\max}(A^TA)}{\lambda_{\min}(A^TA)} = (\mu_2(A))^2$$

Se A è a colonne quasi linearmente dipendenti, $\mu_2(A)$ è grande e $\mu_2(A^TA)$, essendo il suo quadrato, è un grosso amplificatore degli errori sui dati, che sono affetti da errore in quanto provengono da calcoli precedenti. Se poi A è mal condizionata, il calcolo di A^TA può portare a grossi inconvenienti:

$$A = \begin{pmatrix} 1 & 1 \\ \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix} \quad \Rightarrow \quad A^TA = \begin{pmatrix} 1+\varepsilon^2 & 1 \\ 1 & 1+\varepsilon^2 \end{pmatrix} \approx \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

che è singolare se ε^2 è minore della precisione di macchina.

Risoluzione numerica del problema dei minimi quadrati

Un metodo più stabile per il calcolo della soluzione di $A^T A\boldsymbol{\alpha} = A^T \boldsymbol{y}$ è basato sulla decomposizione QR della matrice A mediante trasformazioni elementari ortogonali:

$$A = Q\overline{R} = Q\begin{pmatrix} R \\ \mathbf{0} \end{pmatrix} \Leftrightarrow Q^T A = \begin{pmatrix} R \\ \mathbf{0} \end{pmatrix}$$

con $\overline{R}$ matrice $(m+1) \times (n+1)$, R triangolare superiore di ordine $n+1$ e Q ortogonale di ordine $m+1$. Ricordiamo che si deve risolvere

$$\min_{\boldsymbol{\alpha}} \|A\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2$$

Poiché le trasformazioni ortogonali non alterano la norma euclidea ($\|\boldsymbol{z}\|_2 = \|U\boldsymbol{z}\|_2$ per ogni $\boldsymbol{z} \in R^{m+1}$ e U ortogonale), segue che

$$\|A\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2 = \|Q^T(A\boldsymbol{\alpha} - \boldsymbol{y})\|_2^2 = \|Q^T A\boldsymbol{\alpha} - Q^T\boldsymbol{y}\|_2^2 = \left\| \begin{pmatrix} R \\ \mathbf{0} \end{pmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \tilde{\boldsymbol{y}}_1 \\ \tilde{\boldsymbol{y}}_2 \end{pmatrix} \right\|_2^2$$

dove si è posto $Q^T\boldsymbol{y} = \begin{pmatrix} \tilde{\boldsymbol{y}}_1 \\ \tilde{\boldsymbol{y}}_2 \end{pmatrix} \begin{matrix} \}n+1 \text{ componenti} \\ \}m-n \text{ componenti} \end{matrix}$.

Risoluzione numerica del problema dei minimi quadrati

Allora

$$\|A\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2 = \left\| \begin{pmatrix} R\boldsymbol{\alpha} - \tilde{\boldsymbol{y}}_1 \\ -\tilde{\boldsymbol{y}}_2 \end{pmatrix} \right\|_2^2 = \|R\boldsymbol{\alpha} - \tilde{\boldsymbol{y}}_1\|_2^2 + \|\tilde{\boldsymbol{y}}_2\|_2^2 \geqslant \|\tilde{\boldsymbol{y}}_2\|_2^2$$

dove il segno di uguaglianza vale se e solo se $R\boldsymbol{\alpha} = \tilde{\boldsymbol{y}}_1$.
Il valore minimo possibile vale $\|\tilde{\boldsymbol{y}}_2\|_2^2$ (somma dei quadrati dei residui).
Allora la soluzione del problema si ottiene come soluzione del sistema

$$R\boldsymbol{\alpha} = \tilde{\boldsymbol{y}}_1$$

e deve essere $\boldsymbol{\alpha}^* = R^{-1}\tilde{\boldsymbol{y}}_1$.
La somma dei quadrati dei residui vale $\sum_{i=n+1}^{m} (\tilde{y}_2^{(i)})^2$ e il residuo $\boldsymbol{r} = \boldsymbol{y} - A\boldsymbol{\alpha}$ è tale che

$$\|Q^T(A\boldsymbol{\alpha}^* - \boldsymbol{y})\|_2^2 = \left\| \begin{pmatrix} \mathbf{0} \\ -\tilde{\boldsymbol{y}}_2 \end{pmatrix} \right\|_2^2 \quad \Rightarrow \quad Q^T\boldsymbol{r} = \begin{pmatrix} \mathbf{0} \\ \tilde{\boldsymbol{y}}_2 \end{pmatrix}$$

e pertanto

$$\boldsymbol{r} = Q\begin{pmatrix} \mathbf{0} \\ \tilde{\boldsymbol{y}}_2 \end{pmatrix}.$$

Risoluzione numerica del problema dei minimi quadrati

In tal caso il numero di condizione del problema coincide con il numero di condizione di R:

$$\mu_2(R) = \sqrt{\frac{\lambda_{\max}(R^T R)}{\lambda_{\min}(R^T R)}}$$

Poiché

$$R^T R = (R^T \; 0) \begin{pmatrix} R \\ \mathbf{0} \end{pmatrix} = A^T Q Q^T A = A^T A$$

segue che

$$\mu_2(R) = \sqrt{\frac{\lambda_{\max}(A^T A)}{\lambda_{\min}(A^T A)}} = \mu_2(A)\,.$$

Codice Matlab

```
function [alpha, res, r] = regressLineare(A, y)
%regressLineare - Problema lineare di regressione polinomiale
% INPUT
%  A (double, array) - Matrice di regressione lineare
%  y (double, array) - Vettore delle osservazioni
% OUTPUT
%  alpha (double, array) - Vettore soluzione
%  res   (double)        - Norma 2 del vettore residuo
%  r     (double, array) - Vettore residuo
  [m, n] = size(A);
  [Q, R] = qr(A);
  ytilde = Q' * y;
  alpha = R(1:n, 1:n) \ ytilde(1:n);
  r = Q * [zeros(n,1); ytilde(n+1 : m)]; % residuo: r = y - A*alpha
  res = norm(r,2)^2; % res = norm(ytilde(n+1:m),2)^2;
end
```

L'approssimazione polinomiale mediante un polinomio di grado n relativa a m coppie di dati (x, y) $(m \geqslant n+1)$ si ottiene con

```
>> p = polyfit(x, y, n);
>> z = polyval(p, punti);
```

Se $m = n + 1$ si ottiene interpolazione polinomiale.