

LA RETTA DI REGRESSIONE LINEARE E SISTEMI SOVRADETERMINATI

MAURIZIO PAOLINI - CORSO PAS CLASSE A048

Dipartimento di Matematica e Fisica, Università Cattolica, sede di Brescia. E-mail paolini@dmf.unicatt.it

1. PREMESSA

In queste poche pagine vogliamo affrontare il problema del calcolo della retta di regressione lineare (o retta dei minimi quadrati) collegandolo alla questione dei sistemi lineari sovradeterminati.

Naturalmente il problema può essere affrontato in modo diretto e con un approccio più “statistico”, non c’è nessuna intenzione qui di suggerire quale sia il modo migliore per affrontare il problema.

D’altra parte è opinione dell’autore che affrontare un problema con un’ottica diversa sia sempre positivo e permetta di acquisire maggiore dimestichezza e comprensione.

2. SISTEMI SOVRADETERMINATI

2.1. Notazioni e nozioni di base. Indicheremo con \mathbb{R} l’insieme dei numeri reali, con \mathbb{N} l’insieme dei numeri naturali (zero compreso). Se m e n sono numeri naturali positivi una matrice A $m \times n$ è per noi una “tabella” di numeri reali composta da m righe e n colonne, indicata come

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Indichiamo con $M(m, n)$ l’insieme delle matrici $m \times n$. Si tratta di uno spazio vettoriale su \mathbb{R} di dimensione mn , la somma è definita componente per componente e la moltiplicazione per uno scalare (un numero reale) è anch’essa definita componente per componente. La matrice nulla $0 \in M(m, n)$ ha tutte le componenti nulle.

La “matrice trasposta” $B = A^T$ di una matrice $A \in M(m, n)$ è una matrice $B \in M(n, m)$ definita “scambiando gli indici”, cioè come

$$b_{ij} = a_{ji}$$

Nel caso particolare $n = 1$ abbiamo una matrice con una sola colonna, che identificheremo con lo spazio dei “vettori colonna con m componenti”, indicato con \mathbb{R}^m . Il trasposto di un vettore colonna sarà un “vettore riga” (una matrice $1 \times m$).

Tra una matrice $A \in M(m, p)$ e una matrice $B \in M(p, n)$ è definito il “prodotto righe per colonne” $C = AB \in M(m, n)$ definito da

$$c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

Si noti che **non si può** moltiplicare due matrici se il numero di colonne della prima matrice non è uguale al numero di righe della seconda. Il prodotto righe per colonne **non è** in generale commutativo, ma una semplice verifica permette di dimostrare che è **associativo**, ovvero che $(AB)C = A(BC)$. Un'altra proprietà di immediata verifica coinvolge l'operatore di trasposizione:

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

Se $v, w \in \mathbb{R}^n$, possiamo moltiplicare v^T (matrice $1 \times n$) per w (matrice $n \times 1$) ottenendo una matrice 1×1 , ovvero un numero reale \mathbb{R} ¹. Si ottiene in questo modo il "prodotto scalare" tra vettori, indicato in vari modi:

$$v^T w = v \cdot w = (v, w) \in \mathbb{R}$$

Lo spazio $M(n, n)$ delle matrici "quadrate" $n \times n$ viene per semplicità indicato con la scrittura $M(n)$. Tra due matrici quadrate $n \times n$ il prodotto righe per colonne è ben definito e produce una terza matrice $n \times n$, quindi $M(n)$ è dotato di una operazione di moltiplicazione, oltre che dell'addizione. Con queste due operazioni $M(n)$ verifica le proprietà che ne fanno un "anello". La moltiplicazione **non è commutativa** (salvo il caso particolare $n = 1$), l'elemento neutro della moltiplicazione è la matrice "identità":

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Una matrice quadrata A si dice invertibile o "non singolare" se ammette una inversa moltiplicativa, ovvero una matrice B tale che $AB = BA = I$. Non tutte le matrici sono invertibili, la matrice nulla è un esempio, ma non è l'unica. Le matrici "singolari" sono le matrici non invertibili, un modo per capire se una matrice è singolare consiste nel calcolarne il determinante, se il determinante è nullo la matrice è singolare. Non vogliamo qui insistere sulla definizione e proprietà del determinante di una matrice, per cui rimandiamo ad un testo di base di algebra lineare. Ricordiamo comunque che una matrice quadrata è non singolare se e solo se le sue colonne sono vettori linearmente indipendenti (anche le righe), e quindi formano una base di \mathbb{R}^n .

Definizione 2.1. La matrice $A \in M(n)$ è "simmetrica" se $A = A^T$,

Definizione 2.2. La matrice A è simmetrica semidefinita-positiva se è simmetrica e inoltre

$$x^T A x \geq 0 \quad \text{per ogni } x \in \mathbb{R}^n$$

La matrice A è simmetrica definita-positiva se è simmetrica e inoltre

$$x^T A x > 0 \quad \text{per ogni } x \in \mathbb{R}^n, x \neq 0.$$

Data una matrice $A \in M(m, n)$, un "minore $k \times k$ " è una matrice quadrata ottenuta da A selezionando un sottoinsieme di k righe e un sottoinsieme di k colonne. Naturalmente dovrà essere $k \leq \min\{m, n\}$. Il "minore principale $k \times k$ " è il minore ottenuto selezionando le prime k righe e le prime k colonne.

La matrice A ha rango (almeno) k se esiste almeno un minore $k \times k$ con determinante non nullo.

La matrice A ha rango massimo se ha rango $\min\{m, n\}$ (il valore più alto ammissibile).

¹A rigore bisogna prima identificare lo spazio delle matrici 1×1 che hanno una sola componente con l'insieme dei numeri reali.

2.2. Sistemi lineari e teorema di Rouché-Capelli. Dati $m, n \in \mathbb{N}$ (\mathbb{N} è l'insieme dei numeri naturali, zero compreso, anche se qui richiediamo $n, m > 0$),² la scrittura:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \quad (1)$$

indica un “sistema lineare” di m equazioni in n incognite, con m e n numeri naturali opportuni. Le $a_{ij} \in \mathbb{R}$, $i = 1, \dots, m$, $j = 1, \dots, n$ sono i “coefficienti” del sistema, $b_1, \dots, b_m \in \mathbb{R}$ sono i “termini noti” e $x_1, \dots, x_n \in \mathbb{R}$ sono le quantità incognite. \mathbb{R} indica il campo dei numeri reali. Il sistema (1) si può scrivere in modo più sintetico come $\sum_{j=1}^n a_{ij}x_j = b_i$ per $i = 1, \dots, m$ e utilizzando il “prodotto righe per colonne” si può abbreviare in:

$$Ax = b \quad (2)$$

dove A è la matrice $m \times n$ dei coefficienti, $b \in \mathbb{R}^m$ è il vettore dei termini noti e $x \in \mathbb{R}^n$ il vettore delle incognite. Scritto per esteso, in notazione matriciale, il sistema diventa

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Il “teorema di Rouché-Capelli” permette di stabilire quando il sistema (2) ammette soluzioni (esistenza) quando l'eventuale soluzione è unica (unicità) e quando avvengono entrambe le cose (esistenza ed unicità). Genericamente parlando ci aspettiamo

- infinite soluzioni se $m < n$ (ci sono più incognite che equazioni, sistema sottodeterminato),
- nessuna soluzione se $m > n$ (ci sono più equazioni che incognite, sistema sovradeterminato),
- una e una sola soluzione se $m = n$ (tante equazioni quante incognite),

però, attenzione, questo è vero solo *genericamente parlando*; nello specifico la situazione dipende dai valori effettivi dei coefficienti della matrice e del termine noto, e va precisata in termini di “rango” delle due matrici A e $A|b$ (matrice $m \times (n+1)$ ottenuta da A aggiungendo come ulteriore colonna il vettore b). L'enunciato preciso del teorema di Rouché-Capelli può essere tranquillamente recuperato con una ricerca su internet, ad esempio in wikipedia.

Il caso più importante corrisponde alla scelta $m = n$, in cui c'è esistenza ed unicità della soluzione se e solo se il determinante della matrice A è non nullo, nel qual caso la soluzione può essere scritta in termini della *matrice inversa* A^{-1} come $x = A^{-1}b$.

Noi siamo però interessati al caso di un sistema *sovradeterminato*: $m > n$, in cui generalmente non ci aspettiamo esistenza di soluzioni.

²Se vogliamo proprio fare i “matematici” possiamo intendere che se $m = 0$ si ha un sistema lineare *senza equazioni*, il che significa che tutte le possibili scelte di x_1, \dots, x_n vanno bene, ci saranno infinite soluzioni. Se $n = 0$ non abbiamo incognite, in questo caso le somme a primo membro in (1) si riducono a zero termini, e convenzionalmente si intende che la somma faccia 0 (elemento neutro della somma), il sistema in tal caso ha soluzione (anche se la soluzione è vuota) se e solo se tutti i b_i sono nulli. Se entrambi $m = n = 0$ il sistema lineare ammette convenzionalmente una e una sola soluzione (vuota).

Il *residuo* r di un sistema è definito da

$$r = b - Ax \quad (3)$$

e rappresenta essenzialmente **di quanto** la “soluzione” x **non risolve** il sistema lineare. Ovviamente se $r = 0$, allora x sarà una soluzione, altrimenti non lo sarà.

Nel caso di un sistema sovradeterminato non ci sarà in generale nessun vettore x che annulla il residuo, possiamo però chiederci per quali valori di x il residuo è *il più piccolo possibile*.

2.3. Norme di vettori. L'affermazione *più piccolo possibile* per un vettore pre-suppone che ci sia un modo per quantificare la grandezza di un vettore. Una *norma* è una funzione $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ che, calcolata su un vettore, restituisce un numero reale non negativo e che soddisfa le seguenti proprietà:

- (1) $\|v\| \geq 0$ per ogni vettore v ; (1b) $\|v\| = 0$ se e solo se $v = 0$;
- (2) $\|tv\| = |t|\|v\|$ per ogni vettore v e ogni scalare (numero reale) t ;
- (3) $\|v + w\| \leq \|v\| + \|w\|$ per ogni coppia di vettori v e w (disuguaglianza triangolare).

Ci sono infinite possibili scelte di una norma che soddisfi queste tre richieste; una di queste scelte risulta particolarmente naturale, si tratta della *norma euclidea*, definita da:³

$$\|v\| = \sqrt{\sum_{i=1}^n |v_i|^2}.$$

Denotando con $v \cdot w = \sum_{i=1}^n v_i w_i$ il prodotto scalare tra vettori, la norma euclidea può anche essere scritta come $\|v\|^2 = v \cdot v$. Il prodotto righe per colonne tra un vettore riga e un vettore colonna può essere utilizzato in modo equivalente per indicare il prodotto scalare come $v \cdot w = v^T w$, e quindi

$$\|v\|^2 = v^T v. \quad (4)$$

In assenza di precisazioni assumeremo sempre che $\|\cdot\|$ indichi la norma euclidea, in caso di dubbio la norma euclidea viene indicata con $\|\cdot\|_2$. Altre scelte interessanti di norme sono

- $\|v\|_\infty = \max_{i=1, \dots, n} |v_i|$. Chiamata con vari nomi: norma del massimo, norma “infinito”, norma uniforme, norma cubica, ...
- $\|v\|_1 = \sum_{i=1}^n |v_i|$. Norma “uno”, norma “di Manhattan”, ...

2.4. Soluzione nel senso dei minimi quadrati. Ritornando al sistema lineare sovradeterminato (2) ($m > n$), è utile introdurre una nozione più debole di soluzione.

Definizione 2.3. Diciamo che $x \in \mathbb{R}^n$ *risolve il sistema lineare* $Ax = b$ nel senso dei minimi quadrati se *minimizza la norma euclidea del residuo* r definito da (3).

In altre parole x verifica

$$\|b - Ax\| \leq \|b - Ay\| \quad \text{per ogni } y \in \mathbb{R}^n \quad (5)$$

Nota 2.1. Non è difficile dimostrare che una soluzione nel senso dei minimi quadrati esiste sempre, può tuttavia non essere unica. Si tratta di una nozione più debole di soluzione, infatti se x risolve (2) nel senso usuale, allora si ha $r = b - Ax = 0$, ovvero il residuo ha norma nulla e quindi x risolve (2) anche nel senso dei minimi quadrati. Viceversa è possibile, anzi è normale per sistemi sovradeterminati, che una soluzione nel senso dei minimi quadrati non sia una soluzione nel senso usuale.

³Nel piano euclideo il teorema di Pitagora permette di concludere che la norma euclidea di un “punto del piano”, pensato come vettore con due componenti, non è altro che la sua distanza dall'origine.

Elevando al quadrato la disuguaglianza (5) (si può fare perché abbiamo quantità positive ad ambo i membri) ed utilizzando (4) si ottiene

$$\|b - Ax\|^2 \leq (b - Ay)^T (b - Ay) = y^T A^T Ay - 2b^T Ay + b^T b$$

dove abbiamo usato le proprietà delle matrici: $(Ay)^T = y^T A^T$ e $y^T A^T b = b^T Ay$, l'ultimo passaggio segue perché una matrice 1×1 è sicuramente simmetrica.

In altre parole, x minimizza la funzione

$$f(y) = \frac{1}{2} y^T A^T Ay - b^T Ay$$

e quindi ne annulla il gradiente. Un calcolo (un po' noioso che ci risparmiamo) del gradiente di f porta infine al sistema (detto “**sistema normale**”):

$$A^T Ax = A^T b. \quad (6)$$

Nota 2.2. Se la matrice A , e quindi A^T fosse invertibile (non può esserlo, non essendo quadrata!) il sistema (6) sarebbe equivalente al sistema (2). Per contro la matrice $A^T A$ è una matrice quadrata di dimensione $n \times n$, il suo determinante non è nullo se la matrice A ha rango massimo, ovvero se le sue colonne sono linearmente indipendenti, cosa che assumeremo.

Proposizione 2.1. *La matrice $A^T A$ è simmetrica semidefinita-positiva. Se A ha rango n (rango massimo se $m \geq n$), allora $A^T A$ è definita-positiva.*

Dimostrazione. La simmetria è immediata: $(A^T A)^T = A^T (A^T)^T = A^T A$. Se $v \in \mathbb{R}^n$ è un vettore generico si ha

$$v^T (A^T A)v = (Av)^T (Av) = \|Av\|^2 \geq 0$$

e quindi $A^T A$ è semidefinita-positiva. Per quanto riguarda l'ultima affermazione è sufficiente dimostrare che $A^T A$ è non singolare, in quanto una matrice semidefinita-positiva ma non definita-positiva è necessariamente singolare. Supponiamo quindi per assurdo che esista un vettore $v \in \mathbb{R}^n$ non nullo tale che $(A^T A)v = 0$. In particolare $\|Av\|^2 = v^T A^T Av = 0$ e quindi per le proprietà delle norme deve essere anche $Av = 0$. Il primo membro di quest'ultima equazione può essere interpretata come combinazione lineare delle colonne di A con le componenti di v come coefficienti. Essendo $v \neq 0$ ne segue che le colonne di A sono linearmente dipendenti e quindi A non ha rango n . \square

3. LA RETTA DI REGRESSIONE LINEARE

3.1. Definizione della retta di regressione lineare. Dati m punti del piano, di ascisse x_1, \dots, x_m e ordinate rispettivamente y_1, \dots, y_m , che immaginiamo rappresentino la risposta (variabile y) di un qualche fenomeno sperimentale al variare di un parametro di ingresso (variabile x), ci chiediamo quale sia la retta che meglio li approssima (figura 1). Supporremo per semplicità che le ascisse x_1, \dots, x_m siano tutte distinte. In genere m è piuttosto grande, noi supporremo che sia almeno 3.

Dobbiamo prima di tutto decidere in quale senso si intenda l'approssimazione. La retta cercata avrà equazione

$$y = ax + b$$

e noi immaginiamo che si tratti di una approssimazione (lineare) del fenomeno sperimentale, quindi cercheremo a e b in modo che le quantità $ax_i + b$ approssimino meglio possibile le y_i .

Ovviamente in generale non esisterà una retta che passa esattamente per tutti i punti, ma proviamo ugualmente a scrivere le equazioni che imporrebbero il passaggio della retta per tutti i punti:

$$ax_i + b = y_i, \quad i = 1, \dots, m.$$

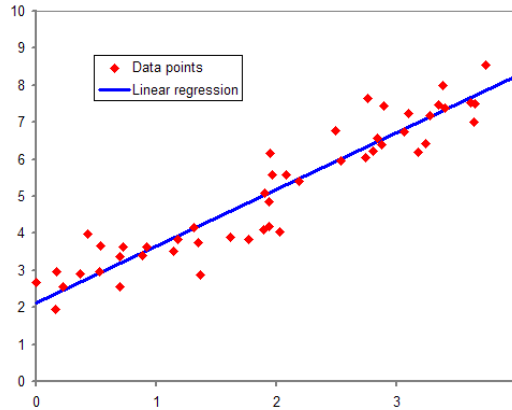


FIGURA 1. La retta di regressione lineare corrispondente ad un insieme di parecchi valori sperimentali.

Qui le incognite sono a e b , mentre tutto il resto è noto; si tratta dunque di un sistema lineare con m equazioni in due incognite, che possiamo scrivere in modo compatto come

$$V\eta = y \quad (7)$$

dove

$$\eta = \begin{bmatrix} b \\ a \end{bmatrix}, \quad V = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

La matrice V è detta matrice di Vandermonde, e in generale ha più di due colonne, ciascuna contenente potenze sempre più alte dei numeri x_1, \dots, x_m . Si tratta di un sistema sovradeterminato con m equazioni e due incognite.

Ci chiediamo: cosa significa risolvere questo sistema *nel senso dei minimi quadrati* (Definizione 2.3)?

Ad ogni punto x_i corrisponde uno scarto $r_i = y_i - ax_i - b$ tra l'ordinata dell' i -esimo punto e l'ordinata della retta nel punto x_i . Curiosamente il vettore $r = [r_1, \dots, r_m]^T$ è proprio il residuo del sistema sovradeterminato (7), quindi la soluzione dei minimi quadrati è quella per cui risulta minima la norma euclidea del vettore degli scarti o equivalentemente il suo quadrato, ovvero la somma dei quadrati degli scarti (è per questo motivo che viene chiamata soluzione *dei minimi quadrati*).

3.2. Calcolo dei coefficienti della retta di regressione lineare. Per risolvere il sistema sovradeterminato (7) nel senso dei minimi quadrati utilizziamo il “sistema normale” (6). La matrice del sistema normale è 2×2 simmetrica, risulta conveniente raccogliere m dalle sue componenti, indicandola quindi con

$$B = V^T V = m \begin{bmatrix} c & d \\ d & e \end{bmatrix}.$$

Calcoliamo separatamente i tre numeri c, d, e . c è il prodotto scalare della prima colonna di V per sé stessa e divisa per m , ovvero

$$c = \frac{1}{m} \sum_{i=1}^m 1 \cdot 1 = 1, \quad d = \frac{1}{m} \sum_{i=1}^m 1 \cdot x_i = \bar{x}, \quad e = \frac{1}{m} \sum_{i=1}^m x_i^2 = \overline{x^2}$$

dove la notazione \bar{x} indica la media aritmetica dei valori x_i e $\overline{x^2}$ la media aritmetica dei quadrati x_i^2 .

Il termine noto del sistema normale, le cui componenti chiameremo mf e mg , è $V^T y$. Si ha

$$f = \frac{1}{m} \sum_{i=1}^m y_i = \bar{y}, \quad g = \frac{1}{m} \sum_{i=1}^m x_i y_i = \overline{xy}$$

dove analogamente a quanto fatto sopra \bar{y} indica la media aritmetica delle y_i e \overline{xy} la media aritmetica dei prodotti $x_i y_i$.

Dividendo ambo i membri per m , il sistema normale si scrive dunque come

$$\begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \overline{xy} \end{bmatrix}$$

Possiamo ricavare il coefficiente angolare a della retta di regressione sottraendo dalla seconda equazione \bar{x} volte la prima equazione:

$$\bar{x}b - \bar{x}b + \overline{x^2}a - \bar{x}^2a = \overline{xy} - \bar{x}\bar{y} \quad \implies \quad a = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}.$$

Pur essendo corretta, conviene riscrivere il valore ottenuto in termini di quantità statistiche come la varianza $\sigma_x^2 = \frac{1}{m} \sum (x_i - \bar{x})^2$ e la covarianza $\sigma_{x,y} = \frac{1}{m} \sum (x_i - \bar{x})(y_i - \bar{y})$. In effetti si ha

$$\begin{aligned} m\sigma_{x,y} &= \sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^m x_i y_i - \bar{y} \sum_{i=1}^m x_i - \bar{x} \sum_{i=1}^m y_i + m\bar{x}\bar{y} = m(\overline{xy} - \bar{x}\bar{y}) \\ m\sigma_x^2 &= \sum_{i=1}^m (x_i - \bar{x})^2 = \sum_{i=1}^m x_i^2 - 2\bar{x} \sum_{i=1}^m x_i + \bar{x}^2 = m(\overline{x^2} - \bar{x}^2) \end{aligned}$$

Quindi in definitiva si può scrivere il coefficiente angolare della retta di regressione come

$$a = \frac{\sigma_{x,y}}{\sigma_x^2}.$$

L'intercetta b della retta di regressione si ottiene poi facilmente usando la prima equazione del sistema normale. Otteniamo:

$$b = \bar{y} - a\bar{x}$$

Nota 3.1. Nell'approccio da noi seguito il fatto che la funzione approssimante sia un polinomio di primo grado non gioca un ruolo essenziale. Si può impostare tranquillamente il problema con $y = ax^2 + bx + c$ (parabola) al posto della retta $y = ax + b$, in altre parole si cerca la parabola (con asse verticale) che meglio approssima i dati nel senso dei minimi quadrati, o anche in generale si può cercare il polinomio di grado n (con n fissato e $n < m$) che minimizza la somma dei quadrati degli scarti. Nel caso generale il sistema 2×2 viene sostituito da un sistema lineare $(n-1) \times (n-1)$. Molti fenomeni (fisici, economici, ecc.) hanno comunque in prima battuta un andamento grossomodo lineare, e quindi il calcolo della retta di regressione è sicuramente il caso più interessante.

Nota 3.2. Osserviamo che nel piano euclideo lo scarto $r_i = y_i - ax_i - b$ viene misurato "in verticale" e quindi **non è** la distanza tra il punto (x_i, y_i) e la retta di regressione. Questo fatto può a prima vista sembrare un punto debole della tecnica di approssimazione, tuttavia:

- Le distanze "verticali" $y_i - ax_i - b$ sono proporzionali alla distanza euclidea dei punti dalla retta (il fattore di proporzionalità dipende dal coefficiente angolare a della retta). Questo tuttavia non implica che la retta di regressione lineare sia anche quella che minimizza la somma dei quadrati anche

nel senso delle distanze “punto-retta”, essendo il coefficiente angolare una delle incognite del problema.

- Nelle applicazioni concrete il significato delle ascisse x_i e delle ordinate y_i è generalmente nettamente diverso, addirittura si tratta spesso di quantità dimensionalmente diverse. Non ha quindi senso misurare le distanze euclidee nel piano x, y , visto che la stessa formula che da la distanza mescola tra loro ascisse e ordinate. In particolare il risultato cambierebbe al variare dell’unità di misura utilizzata per le ascisse x_i o (indipendentemente) per le ordinate y_i .

Nota 3.3. Altre scelte della norma, diverse dalla norma euclidea, portano a problemi diversi. Ad esempio, la norma del massimo $\|\cdot\|_\infty$ porterebbe al seguente problema: *trovare la retta $y = ax + b$ che minimizza il massimo degli scarti $|r_i| = |y_i - ax_i - b|$.* Questo è un problema sensatissimo, forse spesso preferibile all’impostazione nel senso dei minimi quadrati. Tuttavia un problema del genere (di tipo “minimax”) è molto più difficile da studiare e risolvere.