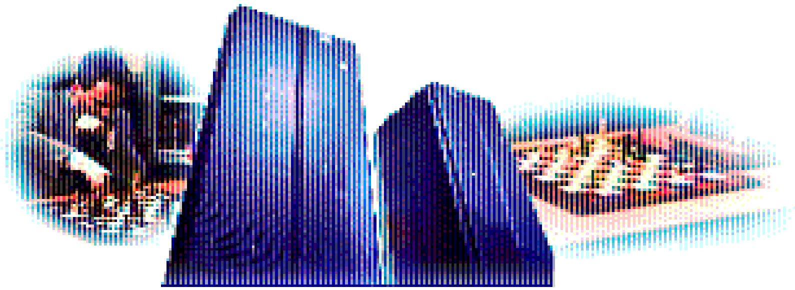
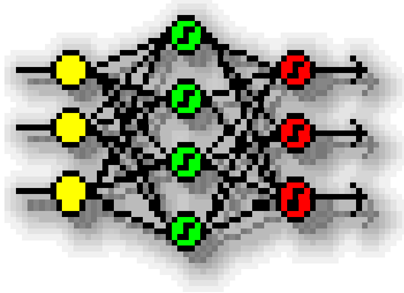




Sistemi Intelligenti
Stimatori e sistemi lineari - III

Alberto Borghese

Università degli Studi di Milano
Laboratory of Applied Intelligent Systems (AIS-Lab)
Dipartimento di Scienze dell'Informazione
borghese@di.unimi.it

Overview

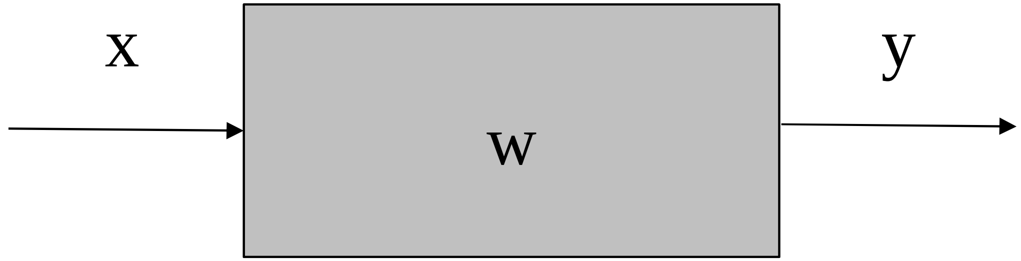
Stima alla massima verosimiglianza di parametri di un modello

Sistemi lineari

Significato geometrico della stima

Relazione tra soluzione di un sistema lineare e stima alla massima verosimiglianza

Modello



$y = f(x \mid w)$

x – causa => y – effetto (misurato)

Identification: determine {w} from {x};{y} - *Supervised learning*

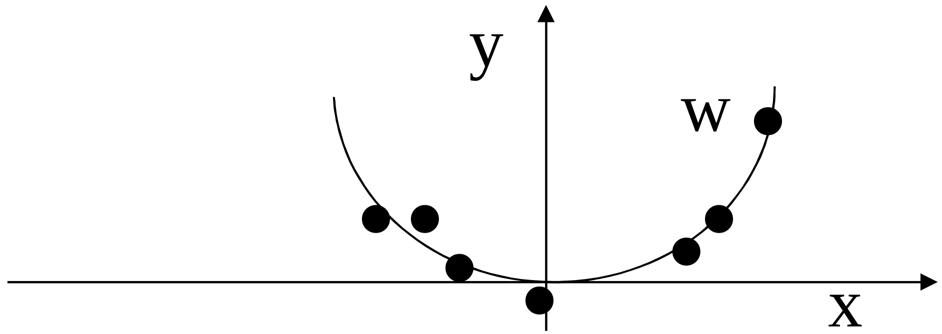
Control / Prediction: determine {y} from {x} {w}

Inverse problem: determine {x} from {y} {w}

Esempio di Modello

$y = f(x; w)$

$y = a\,x^2$



Identificazione: determino i parametri **w** che fittano i punti campionati. La funzione f(x) varie forma con il parametro **a** in questo caso.

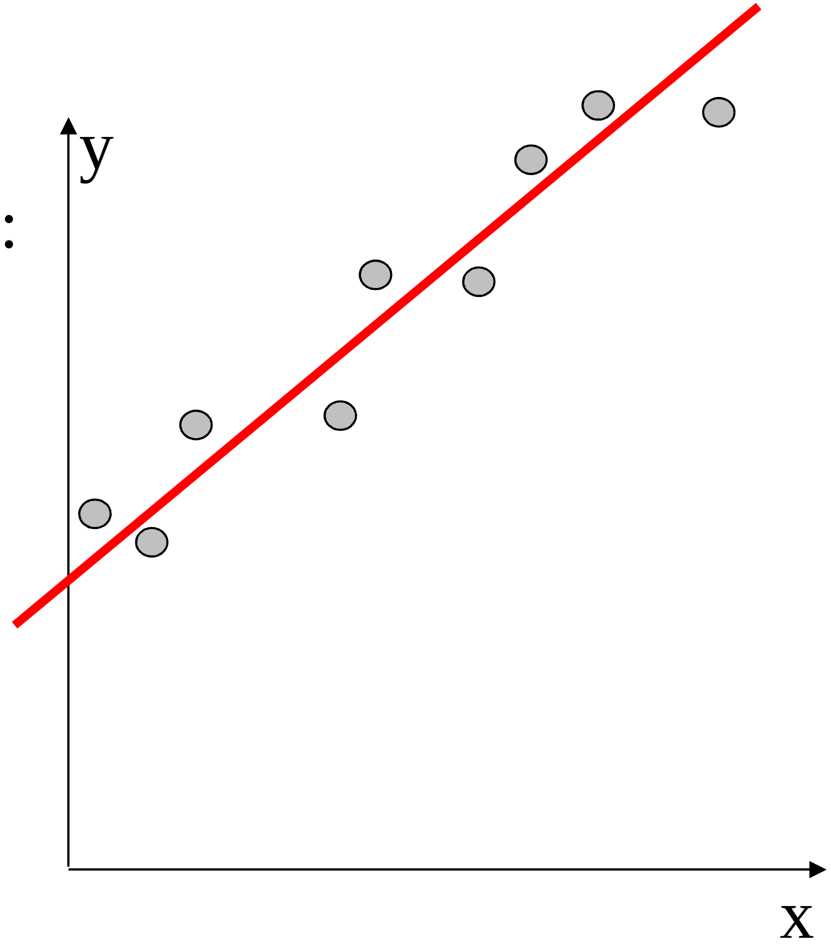
Controllo: Utilizzo il modello (a noto) per predire l'uscita, y_i, in funzione dell'ingresso x_i.

Nota: la funzione f è non lineare in x, ma il modello è lineare in a.

In generale, problemi multi input e multi output: **x, y, w** vettori.

Fitting di una retta

Vogliamo stimare i parametri di una retta: $y = mx + q$, con m e q incogniti: $X = \{m, q\}$



Abbiamo a disposizione N misure effettuate:

$Y = \{y_i ; x_i\}$

Sappiamo che le y_i siano affette da rumore Gaussiano a media nulla. In pratica:

$y_i = y_i + \nu_i$ dove ν_i è l'errore di misura.

Possiamo anche scrivere che:

$y_i = y_{true} + G(\mu, \sigma^2)$ indica una distribuzione monodimensionale gaussiana a media μ e varianza σ^2. Errore di misura: $G(0, \sigma^2)$

Estensione della stima a massima verosimiglianza

$y = f(x \mid w)$ misuro $\{y_i\}$ in corrispondenza di $\{x_i\}$

Avrò che: $f(x_i \mid w) = y_{true} + \nu_i = y_i$

$$L(y \mid x,w) = \prod_i p(v_i) = \prod_i p\left(y_i - f(x_i; w)\right) =$$

$$\prod_i \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{1}{2}\left(\frac{y_i - f(x_i|w)}{\sigma}\right)^2\right)} \right)$$

σ is equal for all samples
$(x_i;y_i)$ are the different samples

Stima ai minimi quadrati e verosimiglianza per modello lineare

- Impostiamo il problema scrivendo la funzione di verosimiglianza e massimizzando tale funzione rispetto a m e q…
- Scriviamo prima di tutto la densità di probabilità di ottenere y_i per ciascun dato:

$$p(x_i, y_i \mid m, q) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2} \left(\frac{y_i - (mx_i + q)}{\sigma} \right)^2 \right]$$

Dove m e q non sono note.

Stima ai minimi quadrati e verosimiglianza

Sapendo che le misure sono indipendenti, possiamo scrivere la probabilità di ottenere le N misure $\{y_i\}$: funzione di verosimiglianza.

Scriviamo il logaritmo negativo della verosimiglianza:

$$f\left(y_1, y_2 .... y_N; m, b; x_1, x_2 ..... x_N\right) = -\sum_{i=1}^{N} \ln\left\{\frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_i - mx_i - b}{\sigma}\right)^2\right]\right\} =$$

$$= -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \sum_{i=1}^{N}\left[-\frac{1}{2}\left(\frac{y_i - mx_i - b}{\sigma}\right)^2\right] =$$

$$= -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - mx_i - b\right)^2$$

Stima ai minimi quadrati e verosimiglianza

- E massimizziamo L(.) ponendo a zero le derivate prime rispetto a m:

$$\frac{\partial f\left(y_1, y_2 .....y_{N;} m, b; x_1, x_2 .....x_N\right)}{\partial m} = \frac{\partial}{\partial m}\left[-\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma}}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right)^2\right] =$$

$$= 0 + \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right)\cdot 2\cdot\left(-x_i\right) =$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right)\cdot x_i = 0 \Rightarrow \sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right)\cdot x_i = 0 \Rightarrow$$

$$\left[\sum_{i=1}^{N}\left(y_i\cdot x_i\right)\right] - m\cdot\left[\sum_{i=1}^{N}\left({x_i}^2\right)\right] - q\cdot\left[\sum_{i=1}^{N}\left(x_i\right)\right] = 0 \Rightarrow$$

$$m\cdot\left[\sum_{i=1}^{N}\left({x_i}^2\right)\right] + q\cdot\left[\sum_{i=1}^{N}\left(x_i\right)\right] = \left[\sum_{i=1}^{N}\left(y_i\cdot x_i\right)\right] \qquad 1^a \text{ equazione}$$

Stima ai minimi quadrati e verosimiglianza

... e a q:

$$= 0 + \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right)\cdot 2\cdot\left(-1\right) =$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right) = 0 \Rightarrow \sum_{i=1}^{N}\left(y_i - m\cdot x_i - q\right) = 0 \Rightarrow$$

$$\left[\sum_{i=1}^{N}\left(y_i\right)\right] - m\cdot\left[\sum_{i=1}^{N}\left(x_i\right)\right] - q\cdot\left[\sum_{i=1}^{N}\left(1\right)\right] = 0 \Rightarrow$$

$$m\cdot\left[\sum_{i=1}^{N}\left(x_i\right)\right] + q\cdot\left[\sum_{i=1}^{N}\left(1\right)\right] = \left[\sum_{i=1}^{N}\left(y_i\right)\right]$$

2ª equazione

Stima massima verosimiglianza

$$\left[\sum_{i=1}^{N}\left(x_i^{\ 2}\right)\right]\cdot m+\left[\sum_{i=1}^{N}\left(x_i\right)\right]\cdot q=\left[\sum_{i=1}^{N}\left(y_i\cdot x_i\right)\right]$$ 1° equazione

$$\left[\sum_{i=1}^{N}\left(x_i\right)\right]\cdot m+\left[\sum_{i=1}^{N}\left(1\right)\right]\cdot q=\left[\sum_{i=1}^{N}\left(y_i\right)\right]$$ 2° equazione

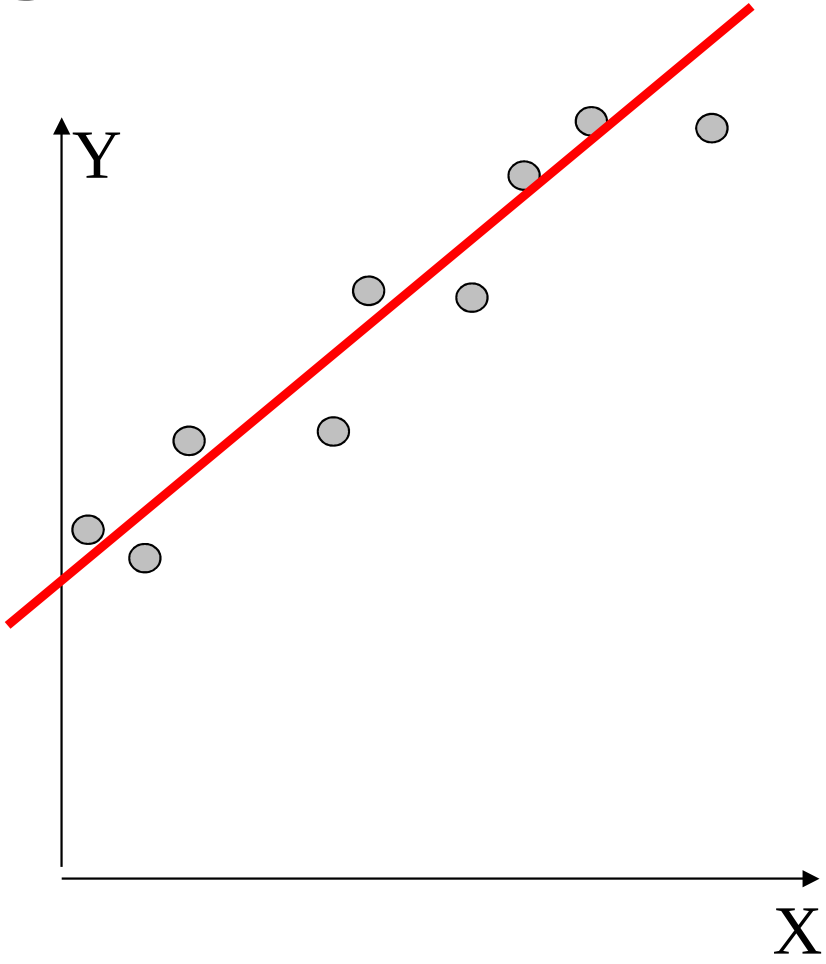
Le incognite, m e b, compaiono con esponente 1 => equazioni lineari in m e q
Potrei risolvere per sostituzione

Stima ai minimi quadrati e verosimiglianza

Per ogni punto, dovrebbe valere
$y_i = mx_i + q$.

Ma c'è l'errore di misura,
misuriamo in realtà $y_i + \nu_i$.

Cerchiamo i parametri m e q
che sono più verosimili.



Cosa vuol dire che sono più verosimili?
Quanto sono più verosimili?

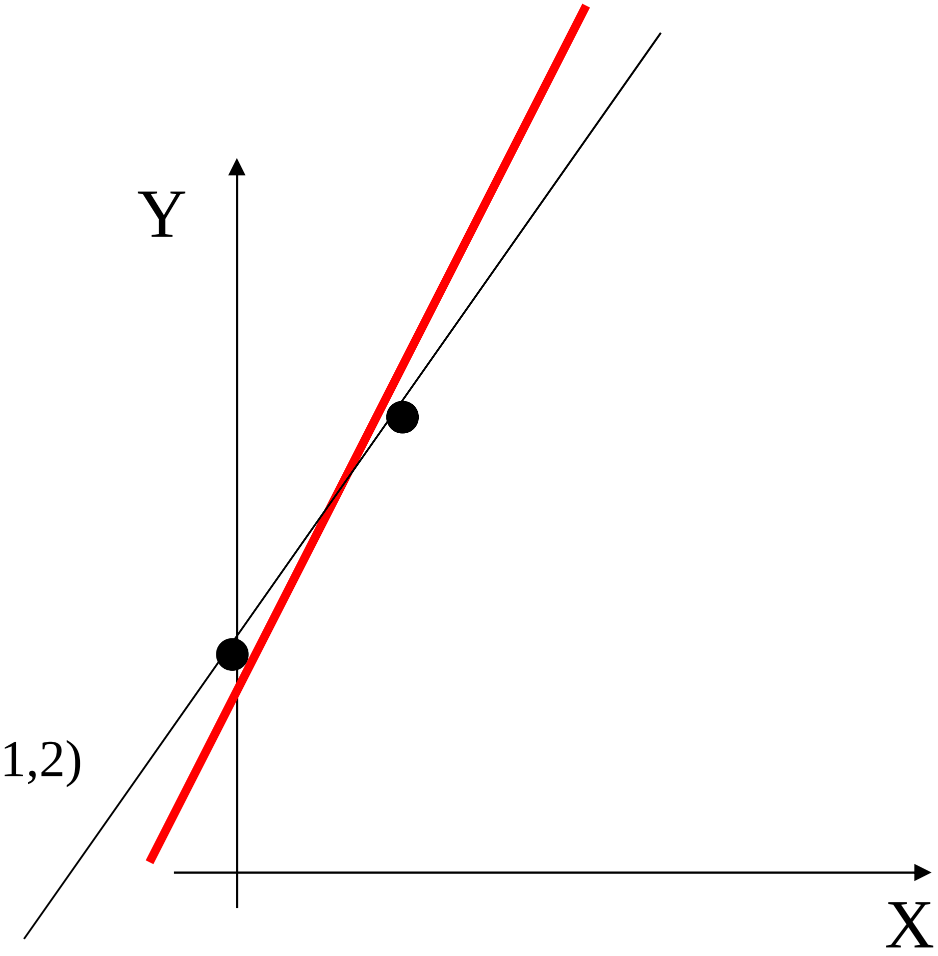
Esempio

$y = 2x + 1$ $\quad\quad$ $m = 2; q = 1$

Misuro e ottengo:

- $x_1 = 1; y_1 = 2{,}8$
- $x_2 = 0; y_2 = 1{,}2$

Quanto varranno le stime di m e q?



$$\left[\sum_{i=1}^{N}\left(x_i^{\,2}\right)\right]\cdot m + \left[\sum_{i=1}^{N}\left(x_i\right)\right]\cdot q = \left[\sum_{i=1}^{N}\left(y_i \cdot x_i\right)\right]$$

$\Rightarrow (1*1 + 0*0)m + (1 + 0)\, q = (1 * 2{,}8 + 0 * 1{,}2)$

$$\left[\sum_{i=1}^{N}\left(x_i\right)\right]\cdot m + \left[\sum_{i=1}^{N}(1)\right]\cdot q = \left[\sum_{i=1}^{N}\left(y_i\right)\right]$$

$\Rightarrow (1 + 0)\, m_e + 2\, q_e = (2{,}8 + 1{,}2)$

$m_e + q_e = 2{,}8 \quad m_e + 2q_e = 4 \quad \Rightarrow$ per sottrazione $\quad q_e = 1{,}2;\ m_e = 1{,}6 \quad$ NB $2{,}8 = 1{,}6*1 + 1{,}2$

Overview

Stima alla massima verosimiglianza di parametri di un modello

Sistemi lineari

Significato geometrico della stima

Relazione tra soluzione di un sistema lineare e stima alla massima verosimiglianza

Sistema lineare

$$a_{11}x_1 + a_{12}\, x_2 + \ldots\ldots. \quad a_{1N}\, x_N = b_1$$
$$a_{21}x_1 + a_{22}\, x_2 + \ldots\ldots. \quad a_{2N}\, x_N = b_2$$

....

$$a_{M1}x_1 + a_{M2}\, x_2 + \ldots\ldots. \quad a_{MN}\, x_N = b_M$$

{aij} – coefficienti in numero N x M

{xj} – incognite, N

{bj} – termini noti, M

I sistemi lineari sono interessanti perchè sono manipolabili con operazioni semplici (algebra delle matrici)

NB le x qui sono i parametri w del modello.

Esempio:

$$3x_1 + 2\, x_2 + \ldots\ldots. \quad 4x_N = 5$$
$$4x_1 -2\, x_2 + \ldots\ldots. \quad 0.5\, x_N = 3$$

....

$$2x_1 + 3\, x_2 + \ldots\ldots. \quad -3\, x_N = -1$$

Matrici

$$A = \left[a_{i,j}\right] \qquad A^T = \left[a_{j,i}\right]$$

$$\alpha A = \left[\alpha a_{i,j}\right] \qquad C = A + B = \left[a_{i,j} + b_{i,j}\right]$$

$$C = AB = \left[c_{i,j}\right] \text{ dove } \left[c_{i,j}\right] = \sum_{k=1}^{n} a_{i,k} b_{k,j}$$

Prodotto degli elementi di una riga per gli elementi di una colonna.

Se A (n x m) ➔ B (m x p) ➔ C (n x p)

$$\mathrm{A} = \begin{bmatrix} 2 & 3 & 1 \\ 1 & -4 & 0 \end{bmatrix} \quad \mathrm{B} = \begin{bmatrix} 1 & -1 \\ 1 & 3 \\ 2 & 0 \end{bmatrix} \implies \mathrm{C} = \begin{bmatrix} 7 & 7 \\ -3 & -13 \end{bmatrix}$$

Se il numero di righe = numero di colonne, matrice quadrata

Matrici (Proprietà)

La somma è associativa e commutativa (A + B) + C = A + (B + C).

Il prodotto è associativo rispetto alla somma ma non gode della proprietà commutativa:

(A+B)C = AC + BC.

AB ≠ BA

$$I = \left[a_{i,j}\right] = \begin{cases} 1 & \text{per } i = j \\ 0 & \text{altrimenti} \end{cases} \quad \text{matrice identità}$$

$$AI = A = IA$$

$$\text{vettore come matrice colonna} : \overline{u}^T = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}$$

$$\text{prodotto vettore matrice} : \overline{v} = \overline{u}^T M$$

Matrice inversa

$$A^{-1}A = I$$

La matrice inversa è definita per una matrice quadrata

Esiste ed è unica se $\det(A) \neq 0$

Numero di condizionamento di una matrice (quadrata): rapporto tra il valore singolare maggiore e minore (cf. Funzione cond in Matlab).
E' una misura di sensibilità della soluzione di un sistema lineare a variazioni nei dati.

Rango di una matrice

Data una matrice A di ordine n (n x n),

una matrice A n x n ha rango m < n se e solo se esiste un suo minore di ordine m non nullo mentre sono nulli tutti i minori di ordine m + 1.

Una matrice A n x n ha rango n (rango pieno) se e solo se il suo determinante è diverso da 0

Rango di una matrice M x N è la dimensione massima di tutte le matrici quadrate estraibili da A e con determinante non nullo. Il rango è massimo quando non è inferiore alla dimensione minima della matrice.

Altre proprietà delle matrici

$\det(AB) = \det(A)\det(B)$

$\det(\mathrm{diag}(W)) = \prod_k w_{k,k}$

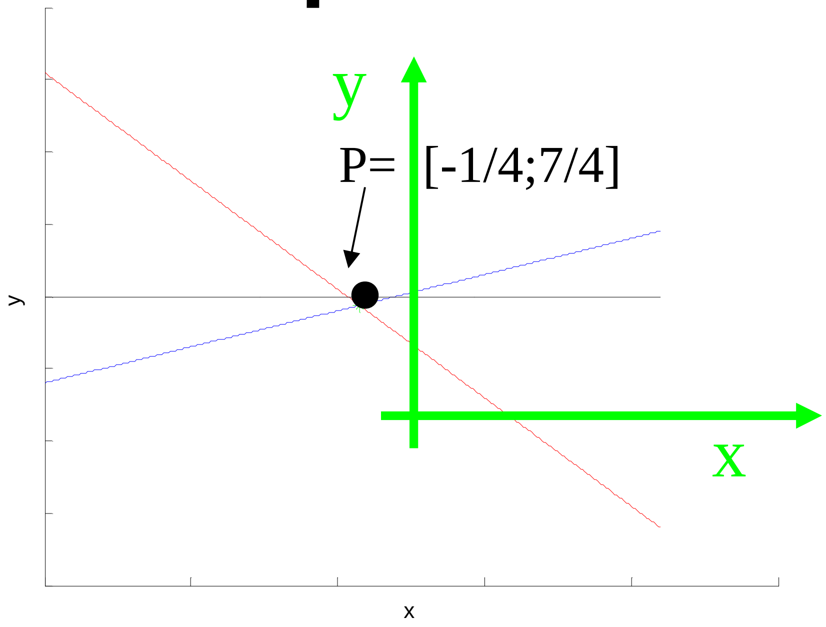
$(A^T)^{-1} = (A^{-1})^T$

$(A\,B\,C)^T = C^T B^T A^T$

Una matrice U, si dice ortogonale se $U^T U = \mathrm{diag}(W)$.
Una matrice U, si dice ortonormale se $U^T U = I \Rightarrow U^{-1} = U^T$

Condizione di ortonormalità:
Il determinante è = 1.
La somma dei prodotti di due righe o di due colonne è = 0.
La somma dei quadrati degli elementi su righe e colonne = 1
Esempio notevole: **matrice di rotazione (cambio di sistema di riferimento)**.

Esempio



$y = x + 2$

$y = -3x + 1$

$1\, x_1 - 1\, x_2 = -2$

$-3\, x_1 - 1\, x_2 = -1$

$y = x_2$

$x = x_1$

Risolvo per sostituzione: $x_1 = -2 + x_2$.

$-3(-2 + x_2) - x_2 = -1 \quad \Rightarrow \quad x_2 = 7/4$

$x_1 - 1/4 = 2 \quad \Rightarrow \quad x_1 = -1/4$

Sistema lineare

$a_{11}x_1 + a_{12}\, x_2 + \ldots\ldots\ \ a_{1N}\, x_N = b_1$

$a_{21}x_1 + a_{22}\, x_2 + \ldots\ldots\ \ a_{2N}\, x_N = b_2$

....

$a_{M1}x_1 + a_{M2}\, x_2 + \ldots\ldots\ \ a_{MN}\, x_N = b_M$

Esempio:

$3x_1 + 2\, x_2 + \ldots\ldots\ \ 4x_N = 5$

$4x_1 - 2\, x_2 + \ldots\ldots\ \ 0.5\, x_N = 3$

....

$2x_1 + 3\, x_2 + \ldots\ldots\ \ -3\, x_N = -1$

$\mathbf{A}\, \mathbf{x} = \mathbf{b}$

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

M x 1

Vettore dei termini noti

Sistema quadrato (N x N)

$a_{11}x_1 + a_{12}\, x_2 + \ldots\ldots\quad a_{1N}\, x_N = b_1$

$a_{21}x_1 + a_{22}\, x_2 + \ldots\ldots\quad a_{2N}\, x_N = b_2$

....

$a_{N1}x_1 + a_{N2}\, x_2 + \ldots\ldots\quad a_{NN}\, x_N = b_N$

Esempio:

$3x_1 + 2\, x_2 + \ldots\ldots\quad 4x_N = 5$

$4x_1 -2\, x_2 + \ldots\ldots\quad 0.5\, x_N = 3$

....

$2x_1 + 3\, x_2 + \ldots\ldots\quad -3\, x_N = -1$

Ammette 1, nessuna o ∞ soluzioni

A è N x N quadrata

$\mathbf{A\, x = b}$

$\mathbf{A^{-1}Ax = A^{-1}b}$

$\mathbf{x} = \mathbf{A^{-1}\, b}$ se $\mathbf{A^{-1}}$ esiste, 1 soluzione.

altrimenti, nessuna (rette parallele)

o

∞ soluzioni (rette coincidenti).

Soluzione dei sistemi lineari

$y = x + 2$

$y = -3x + 1$

Scrivo il sistema lineare: $Ax = b$

$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

$$1\,x_1 - 1\,x_2 = -2$$
$$-3\,x_1 - 1\,x_2 = -1$$

X è una soluzione se soddisfa tutte le equazioni del sistema stesso.

Soluzioni:

- ! ∃ Soluzione (sistema impossibile)
- ∃ Soluzione (sistema possibile)
  - 1 soluzione (sistema determinato)
  - > 1 soluzione (∞^k soluzioni – sistema indeterminato).

Soluzione di sistemi lineari quadrati

$$x = A^{-1} b$$

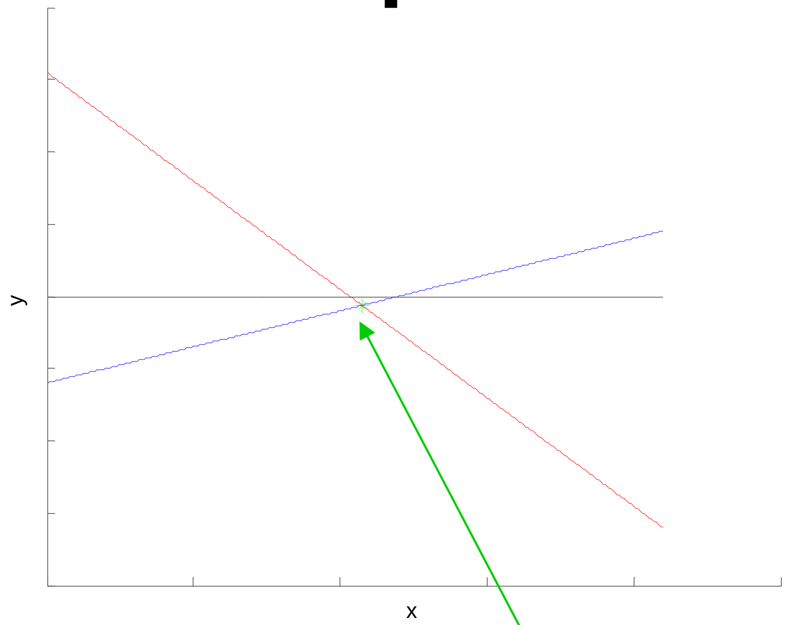
Condizione di esistenza dell'inversa è $\det(A) \neq 0$

Il sistema ammette 1 ed 1 sola soluzione se $\det(A) \neq 0$

Altrimenti: nessuna o infinite soluzioni

Esempio

$$y = x + 2$$
$$y = -3x + 1$$



$$1\,x_1 - 1\,x_2 = -2$$
$$-3\,x_1 - 1\,x_2 = -1$$

$$x_1 = x$$
$$x_2 = y$$

$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

Rango di A è pieno

$$\det(A) = 1(-1)-(-1)(-3)=-1-3=-4$$

$$x_1 = -1/4$$
$$x_2 = 7/4$$

$$P = A^{-1}\,b$$

P = [-1/4 7/4]

Risoluzione di un sistema 2x2

$a_{11}x_1+a_{12}x_2 = b_1$
$a_{21}x_1+a_{22}x_2 = b_2$

$\mathbf{y} = \mathbf{Ax}$

$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{y}$

$$A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

$$\det(A) = a_{11}*a_{22}-a_{12}*a_{21}$$

Esempio di soluzione non univoca

$y = x + 2$

$2y = 2x + 3$

$1\,x_1 - 1\,x_2 = -2$

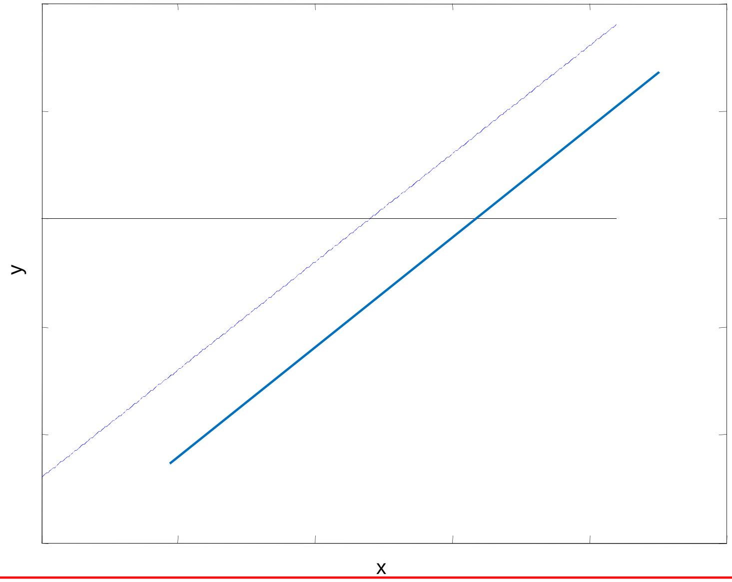
$2\,x_1 - 2\,x_2 = -3$

$x_1 = x$

$x_2 = y$

$$A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ -3 \end{bmatrix}$$

Non esistono soluzioni



det(A) = 1(-2)-(-1)(2)=-2+2=0 La soluzione non esiste o ∞ soluzioni.

$y = x + 2$

$2y = 2x + 4$

La soluzione, se esiste non è unica: tutti i punti della retta soddisfano contemporaneamente le 2 equazioni. In questo caso ∞ soluzioni: rette sovrapposte.

Sistema M x N, M > N

$a_{11}x_1 + a_{12}\ x_2 + ...\quad a_{1N}\ x_N = b_1$

$a_{21}x_1 + a_{22}\ x_2 + ...\quad a_{2N}\ x_N = b_2$

....

$a_{M1}x_1 + a_{M2}\ x_2 + ...\quad a_{MN}\ x_N = b_M$

Esempio:

$3x_1 + 2\ x_2 + ...\quad 4x_N = 5$

$4x_1 -2\ x_2 + ...\quad 0.5\ x_N = 3$

....

$2x_1 + 3\ x_2 + ...\quad -3\ x_N = -1$

Ammette 1, nessuna o ∞ soluzioni

A x = b

A è M x N, M > N, non è una matrice quadrata.

1, nessuna, ∞ soluzioni.

Ho delle equazioni di troppo, devono essere correlate (combinate linearmente), perché il sistema ammetta soluzione.

Posso sempre calcolare la soluzione in forma matriciale.

Sistemi lineari con m > n

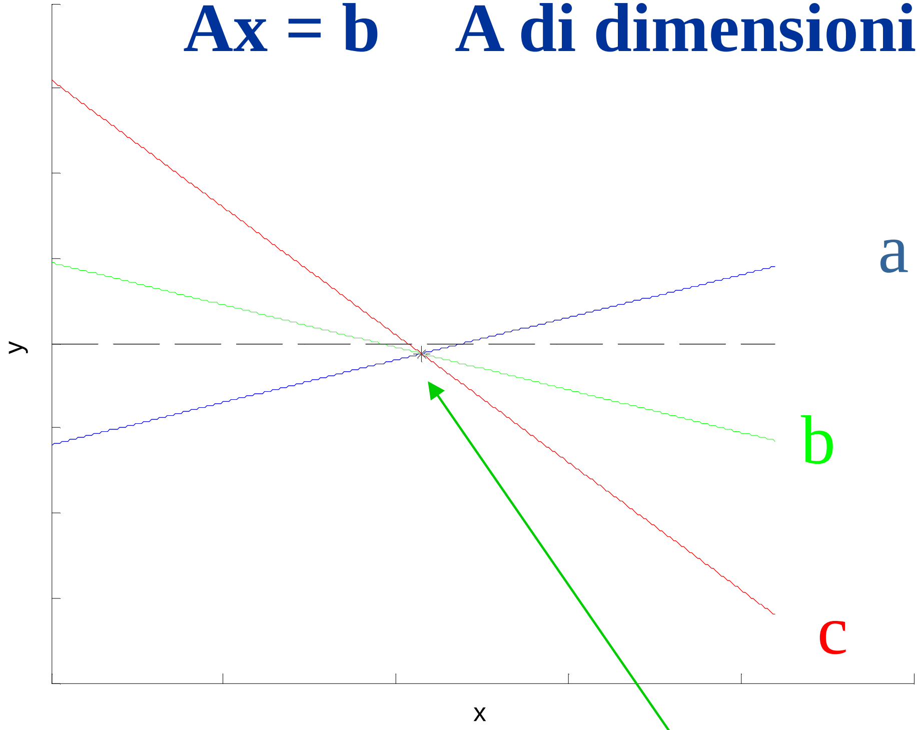
J(W,L) è rettangolare: numero di righe maggiore del numero di colonne

Ax = b   A di dimensioni m x n

y = x + 2

y = -3x + 1

y = -x + 3/2



Una delle 3 righe di A è combinazione lineare delle altre.

$$A=\begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \\ -1.5 \end{bmatrix}$$

Esiste un'equazione "di troppo"

Nessuna, 1 o ∞ soluzioni

$\boxed{\text{Rango di A è pieno}}$

Rango di una matrice

det ($A^{\S}ij$) Minore complementare

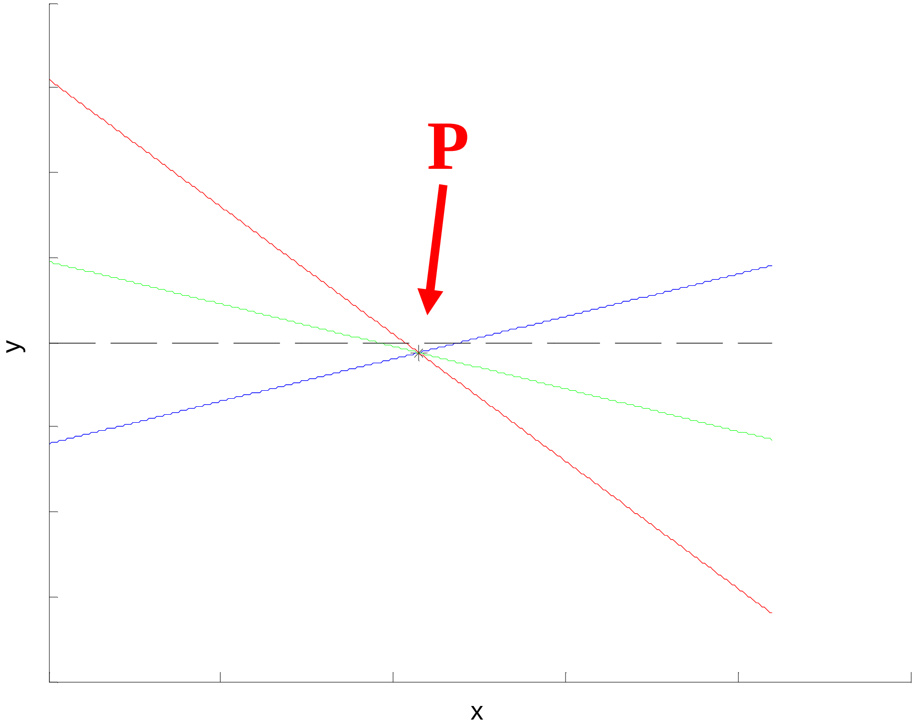
Data una matrice A di ordine n (n x n),

una matrice A n x n ha rango m < n se e solo se esiste un suo minore di ordine m non nullo

mentre sono nulli tutti i minori di ordine m + 1.

Una matrice A n x n ha rango n (rango pieno) se e solo se il suo determinante è diverso da 0

Relazione tra le equazioni (combinazione lineare)



$\alpha_1 (y - x - 2) +$

$\alpha_2 (y + 3x - 1) =$

$(y + x - 3/2)$

In questo caso:

$\alpha_1 = -1/2$

$\alpha_2 = -1/2$

Tutte le rette per la soluzione P possono essere descritte come un fascio (di rette).

Un fascio di rette è univocamente identificato da due rette (che si incontrino in un punto).

La terza equazione è combinazione lineare delle prime due.

Sistema lineare: soluzione algebrica

Caso generale:

$\mathbf{A\,x = b} \Rightarrow \mathbf{A'\,A\,x = A'\,b} \Rightarrow \mathbf{(A'\,A)^{-1}A'\,A\,x = (A'\,A)^{-1}A'\,b}$

$\Downarrow$

(A'A) gioca il ruolo di A quadrata.

$\mathbf{x = (A'A)^{-1}A'b}$

Quale criterio viene soddisfatto da **x**?

$C = (A^T * A)^{-1}$ è la matrice di covarianza (matrice quadrata n x n)

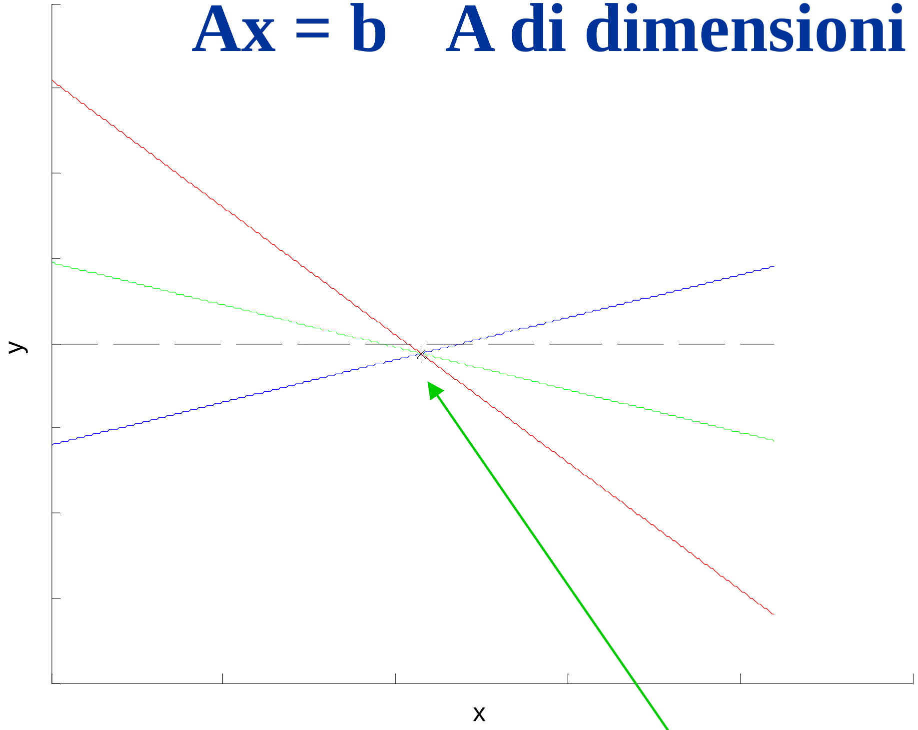
Sistemi lineari con m > n

Ax = b   A di dimensioni m x n

$y = x - 2$

$y = -3x + 1$

$y = -x + 3/2$



P = [-1/4 7/4]

$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \\ +1.5 \end{bmatrix}$$

$$A^T * A = \begin{bmatrix} 11 & 3 \\ 3 & 3 \end{bmatrix} \qquad \det = 24$$

$$C = (A^TA)^{-1} = \begin{bmatrix} 0.1250 & -0.1250 \\ -0.1250 & 0.4583 \end{bmatrix}$$

$$P = C * A^T * b$$

P = [-0.25 +1.75]

intersezione

Riformulazione del problema con rumore

$a_{11}x_1 + a_{12}\, x_2 + \ldots\ldots \quad a_{1N}\, x_N = b_1 + \nu_1$

$a_{21}x_1 + a_{22}\, x_2 + \ldots\ldots \quad a_{2N}\, x_N = b_2 + \nu_2$

....

$a_{M1}x_1 + a_{M2}\, x_2 + \ldots\ldots \quad a_{MN}\, x_N = b_M + \nu_M$

Modello **Misure**

Errore di modello (sistematico, randomico). M x 1 => **Residuo.**

$$\mathbf{A}\,\mathbf{x} = \mathbf{b} + \mathbf{N}$$

M x 1

Vettore dei termini noti

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

Quale criterio viene soddisfatto da **x**?

Soluzione come problema di ottimizzazione

Funzione costo: $(Ax - b)^2 = \sum_k \nu_k{}^2 = \|Ax - b\|^2$

Assegno un costo al fatto che la soluzione x, non soddisfi tutte le equazioni, la somma dei residui associati ad ogni equazioni viene minimizzata. Geometricamente: viene trovato il punto a distanza (verticale) minima da tutte le rette.

$$\min_{\mathbf{x}} \sum_k \nu_k{}^2 = \min_{\mathbf{x}} (\mathbf{Ax} - \mathbf{b})^2$$

$$\frac{d}{d\mathbf{x}}(\mathbf{Ax} - \mathbf{b})^2 = 2\mathbf{A}^{\mathbf{T}}(\mathbf{Ax} - \mathbf{b}) = 0$$

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

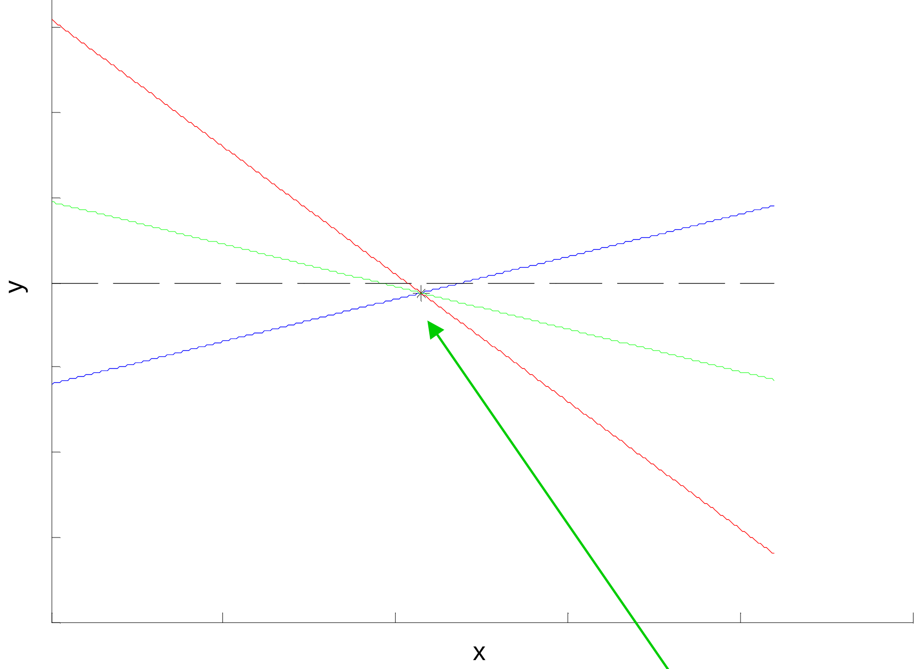
$$\mathbf{x = (A^T A)^{-1} A^T b}$$

NB le funzioni costo sono spesso quadratiche (problemi di minimizzazione convessi) perchè il costo cresce sia che il modello sovrastimi che sottostimi le misure.
Inoltre, le derivate calcolate per imporre le condizioni di stazionarietà (minimo), sono relativamente semplici.

Sistemi lineari con m > n

Ax = b A di dimensioni m x n

$$y = x - 2$$
$$y = -3x + 1$$
$$y = -x + 3/2$$



P = [-1/4 7/4]

$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \\ -1.5 \end{bmatrix}$$

$$A^T * A = \begin{bmatrix} 11 & 3 \\ 3 & 3 \end{bmatrix} \qquad \det = 24$$

$$C = (A^T A)^{-1} = \begin{bmatrix} 0.1250 & -0.1250 \\ -0.1250 & 0.4583 \end{bmatrix}$$

$$P = C * A^T * b$$

P = [-0.25 +1.75]

intersezione

$$||Ax - b|| = 0$$

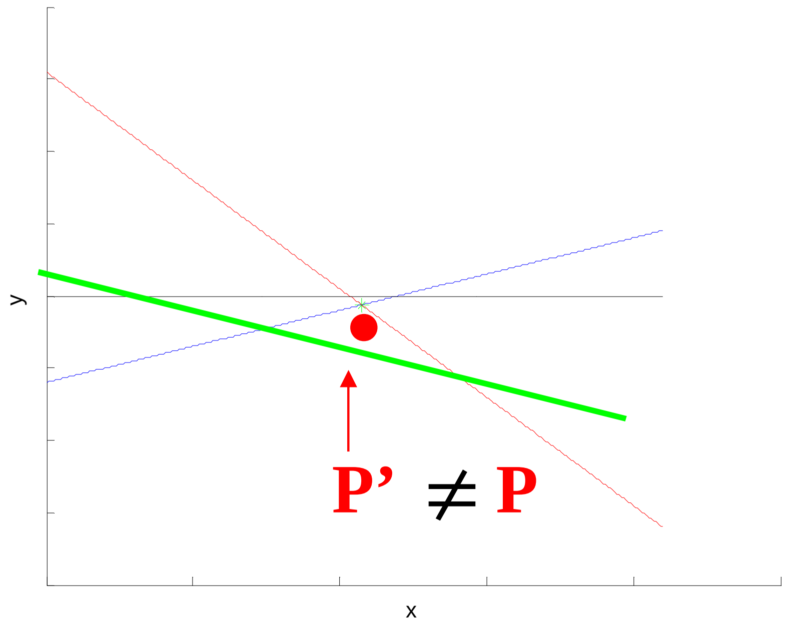
Sistemi lineari con m > n - non esiste soluzione (matematica)

AX = b **A di dimensioni m x n**

y = x + 2

y = -3x + 1

y = -x + 1/2



$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \\ -0.5 \end{bmatrix}$$

$$A^T * A = \begin{bmatrix} 11 & 3 \\ 3 & 3 \end{bmatrix} \qquad \det = 24$$

$$\sum_k \nu_k^{\,2} = \|Ax - b\|^2 = 0.333333$$

$$C = (A^TA)^{-1} = \begin{bmatrix} 0.1250 & -0.1250 \\ -0.1250 & 0.4583 \end{bmatrix}$$

$$P = C * A^T * b$$

P' = [-0.5 +1.4167]

No intersezione

Commenti

$$\sum_k {\nu_k}^2 = \|Ax - b\|^2 = \sum_k \| A_{k,*}x - b_k \|^2 =$$

$$\left[\left(A_{11}x_1 + A_{12}x_2\right) - b_1\right]^2 + \left[\left(A_{21}x_1 + A_{22}x_2\right) - b_2\right]^2 +$$

$$\left[\left(A_{31}x_1 + A_{32}x_2\right) - b_3\right]^2$$

Lo scarto misura la distanza (verticale) dalla retta

Overview

Stima alla massima verosimiglianza di parametri di un modello

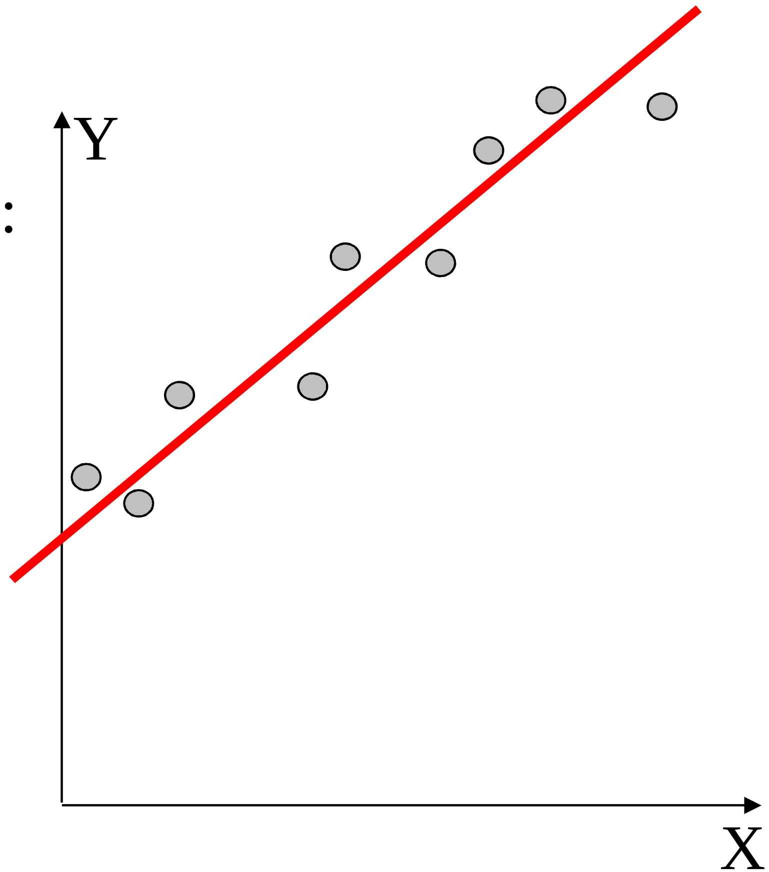
Sistemi lineari

Significato geometrico della stima

Relazione tra soluzione di un sistema lineare e stima alla massima verosimiglianza

Fitting di una retta

Vogliamo stimare i parametri di una retta: y = mz + q, con m e q incogniti:
X = {m, q}



Abbiamo a disposizione N misure effettuate:

$Y = \{y_i ; x_i\}$

Sappiamo che le y_i sono affette da rumore Gaussiano a media nulla. In pratica:

$y_i = y_i + \nu_i$ dove ν_i è il rumore di misura.

Possiamo anche scrivere che:

$y_i = G(mx_i + b, \sigma^2)$, dove $G(\mu, \sigma^2)$ indica una distribuzione monodimensionale gaussiana a media μ e varianza σ^2.

Stima alla massima verosimiglianza

- Impostiamo il problema scrivendo la funzione di verosimiglianza e massimizzando tale funzione rispetto a m e q…
- Scriviamo prima di tutto la densità di probabilità di ottenere y_i per ciascun dato:

$$p(y_i \mid m, q; x_i) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_i - mx_i - q}{\sigma} \right)^2 \right]$$

Stima alla massima verosimiglianza

Scriviamo il logaritmo negativo della verosimiglianza:

$$f\left(y_1, y_2 \ldots . y_N; m, b; x_1, x_2 \ldots .. x_N\right) = -\sum_{i=1}^{N} \ln\left\{\frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_i - mx_i - q}{\sigma}\right)^2\right]\right\} =$$

$$= -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \sum_{i=1}^{N}\left[-\frac{1}{2}\left(\frac{y_i - mx_i - q}{\sigma}\right)^2\right] =$$

$$= -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - mx_i - q\right)^2$$

Minimizzo f(.) rispetto a {m.q}

Stima massima verosimiglianza

$$\left[\sum_{i=1}^{N}\left(x_i^{\,2}\right)\right]\cdot m+\left[\sum_{i=1}^{N}\left(x_i\right)\right]\cdot b=\left[\sum_{i=1}^{N}\left(y_i\cdot x_i\right)\right]$$

1° equazione

$$\left[\sum_{i=1}^{N}\left(x_i\right)\right]\cdot m+\left[\sum_{i=1}^{N}\left(1\right)\right]\cdot b=\left[\sum_{i=1}^{N}\left(y_i\right)\right]$$

2° equazione

Le incognite, m e b, compaiono con esponente 1 => equazioni lineari in m e b: X={m,q}.

Posso risolvere come:

$$\mathbf{A^TAx = A^Tb}$$

$$\mathbf{x} = \mathbf{(A^TA)^{-1}A^Tb}$$

Stima alla massima verosimiglianza: caso 2D

- Ottengo un sistema di 2 equazioni in 2 incognite

$$\begin{bmatrix} \sum_{i=1}^{N}\left(x_i^{\,2}\right) & \sum_{i=1}^{N}\left(x_i\right) \\ \sum_{i=1}^{N}\left(x_i\right) & \sum_{i=1}^{N}\left(1\right) \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N}\left(y_i \cdot x_i\right) \\ \sum_{i=1}^{N}\left(y_i\right) \end{bmatrix}$$

A x = b →

X = (A'A)$^{-1}$A'B = CA'B - C è matrice di covarianza.

NB Le y_i sono solo ai termini noti

Stima ai minimi quadrati caso 2D

- Scriviamo l'equazione della retta per tutti i punti in forma matriciale (sistema lineare Ax=b, N equazioni, 2 incognite):

$$\begin{bmatrix} x1 & 1 \\ x2 & 1 \\ ... & ... \\ xN & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y1 \\ y2 \\ ... \\ yN \end{bmatrix}$$

- Vogliamo trovare x t.c. $(Ax-b)^T(Ax-b)$ è minima (minimizzazione dei quadrati dei residui).

X = **(A'A)$^{-1}$A'B = CA'B** - C è matrice di covarianza.

Esempio - Caso 2D (2 parametri)

N = 20 punti

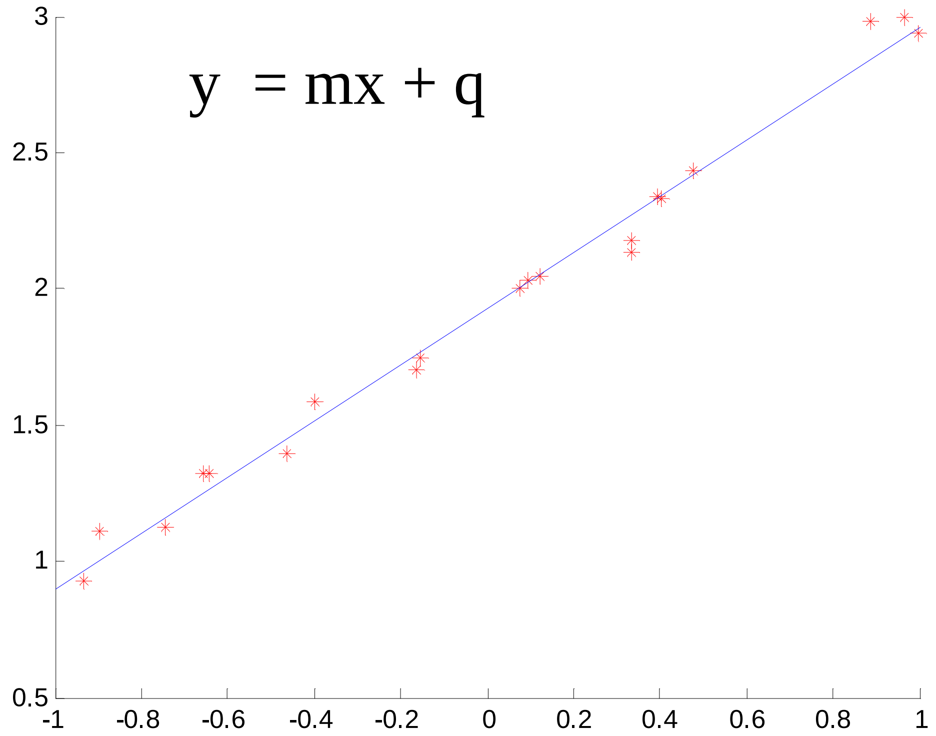
$\sigma_o^2 = 0.01$

m reale = 1

q reale = 2

m stimato = 0.9931

q stimato = 2.0106



Cosa vuol dire che {m,q} sono i più verosimili?
Quanto sono più verosimili?

Stima a massima verosimiglianza e minimi quadrati

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

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

La soluzione a massima verosimiglianza, quando il rumore è Gaussiano a media nulla, coincide con la soluzione ai minimi quadrati del sistema lineare associato (la soluzione ai minimi quadrati è un caso particolare della stima alla massima verosimiglianza).

La soluzione è quella che minimizza lo scarto quadratico medio dei residui, ovverosia è a minima varianza.

La stima a massima verosimiglianza è un approccio generale, e si presta a p(x) di qualsiasi forma. La Gaussiana consente di ottenere una formulazione lineare del problema.

Stima ai minimi quadrati e verosimiglianza

- Nella soluzione ai minimi quadrati del sistema lineare $Ax=b$ si definisce un vettore errore $v = Ax - b$;
- Nel caso di soluzione “perfetta” $v = 0$;
- Dal momento che abbiamo un numero di equazioni maggiore rispetto al numero di incognite, cerchiamo il vettore e a norma minima;
- In pratica cerchiamo x t.c. $v^T v = \Sigma_i v_i^2$ è minimo.

Giustificazione statistica

- **C'è un solo insieme vero dei parametri**, mentre ci **possono essere infiniti universi di dati per effetto dell'errore di misura**.

- La domanda quindi più corretta sarebbe: "Dato un certo insieme di parametri, qual'è la probabilità che questo insieme di dati sia estratto?" (più correttamente si parla di densità di probabilità?)

- Cioè, **per ogni insieme di parametri, calcoliamo la probabilità che i dati siano estratti. Ovverosia la likelihood (verosimiglianza) dei dati, dato un certo insieme di parametri.**

La stima ai minimi quadrati dei parametri è equivalente a determinare i parametri che massimizzano la funzione di **verosimiglianza** sotto l'ipotesi di errore **Gaussiano a media nulla**.

Overview

Stima alla massima verosimiglianza di parametri di un modello

Sistemi lineari

Significato geometrico della stima

Relazione tra soluzione di un sistema lineare e stima alla massima verosimiglianza