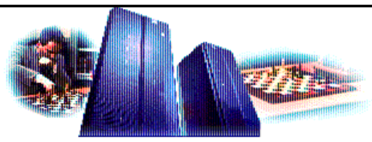
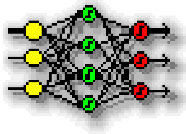




Sistemi Intelligenti
Stimatori e sistemi lineari - II

Alberto Borghese

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

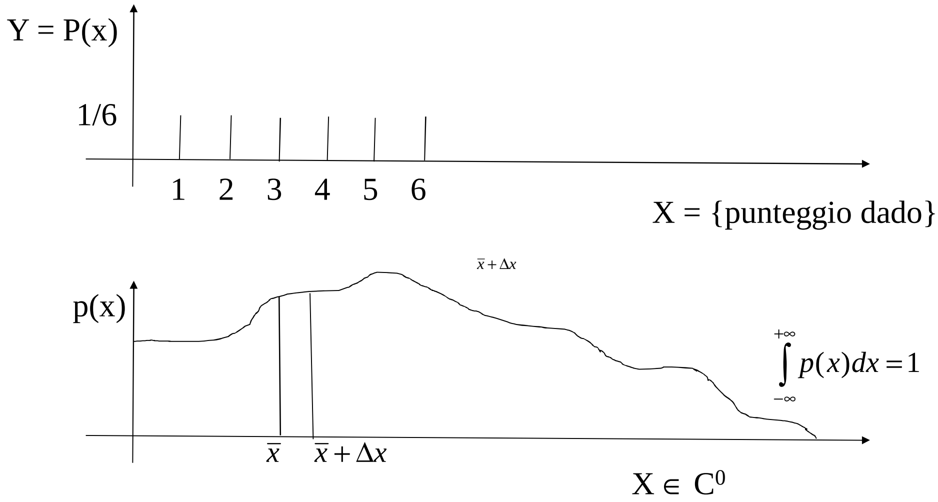
Overview

Distribuzioni di probabilità

Stima alla massima verosimiglianza

Sistemi lineari

La probabilità nel caso continuo



$$P(x \in [\bar{x}, \bar{x}+\Delta x]) \int_{\bar{x}}^{\bar{x}+\Delta x} p(x)dx$$

Definizione di p(x)

Caso discreto: prescrizione della probabilità per ognuno dei finiti valori che la variabile X può assumere: P(X).

Caso continuo: i valori che X può assumere sono infiniti. Devo trovare un modo per definirne la probabilità. Descrizione **analitica** mediante la funzione densità di probabilità.

Valgono le stesse relazioni del caso discreto, dove alla somma si sostituisce l'integrale.

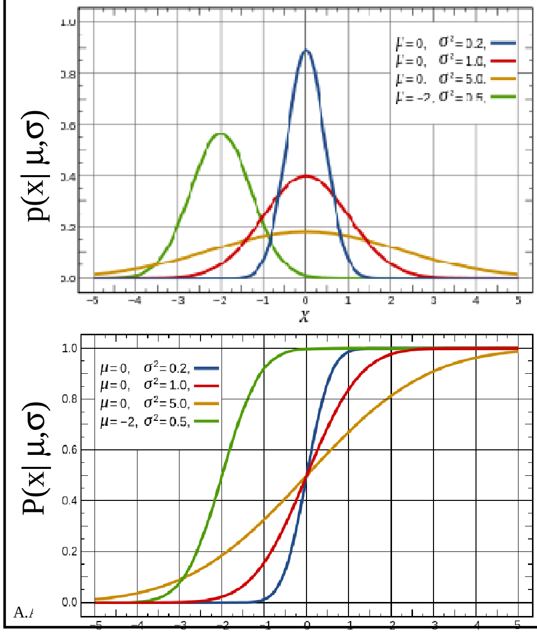
$$P(X = x \in [\bar{x}, \bar{x}+\Delta x]) \int_{\bar{x}}^{\bar{x}+\Delta x} \int_{-\infty}^{+\infty} p(x,y)dxdy$$

$$p(x,y) = p(y|x)\, p(x) = p(x \mid y)\, p(y)$$

$$p(x \mid y) = \frac{p(y|x)\, p(x)}{p(y)}$$

Come esprimo la forma analitica di p(x)?

Distribuzioni notevoli: la Gaussiana



$$p(x|\mu,\sigma)=\frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^D}\cdot\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

D = dimensione, in questo caso D = 1

$Pr(| X-\mu | < \sigma) = 0.68268$

$Pr(| X-\mu | < 2\sigma) = 0.95452$

$Pr(| X-\mu | < 3\sigma) = 0.9973$

I momenti di una variabile statistica

$$\mu^k(X)=\int_{-\infty}^{+\infty}(x-a)^k\ p(x)dx$$

Momento rispetto ad a, solitamente alla media

$$E[X]=\int_{-\infty}^{+\infty}x\ p(x)$$

Valore atteso (Expeted value) di X = media distribuzione

$$E\left[(X-\mu)^2\right]=\int_{-\infty}^{+\infty}(x-\mu)^2\ p(x)$$

Varianza (σ^2)

$$E\left[(X-\mu)^3\right]=\int_{-\infty}^{+\infty}(x-\mu)^3\ p(x)$$

Asimmetria

$$E\left[(X-\mu)^4\right]=\int_{-\infty}^{+\infty}(x-\mu)^4\ p(x)$$

Kurtosi – peso delle code di p(x)

Overview

Distribuzioni di probabilità

Stima alla massima verosimiglianza

Sistemi lineari

Nozioni di base

Variabili indipendenti: $p(y_1,y_2) = p(y_1)p(y_2)$

Gaussiana: siano date due realizzazioni **indipendenti** della stessa variabile casuale x… Quale è la probabilità di misurare y_1 nella prima realizzazione e y_2 nella seconda realizzazione (estrazione con ripetizione)?

$$p(y_1 \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_1-\mu}{\sigma}\right)^2\right]$$

$$p(y_2 \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_2-\mu}{\sigma}\right)^2\right]$$

$$p(y_1, y_2 \mid \mu,\sigma) = p(y_1 \mid \mu,\sigma) \cdot p(y_2 \mid \mu,\sigma) =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_1-\mu}{\sigma}\right)^2\right] \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_2-\mu}{\sigma}\right)^2\right]$$

Funzione di verosimiglianza

- Siano date **N variabili casuali <u>indipendenti</u>**… Quale è la **probabilità di misurare il vettore [y_1, …, y_N]**?

$$p(y_1, y_2, ..., y_N) = p(y_1) \cdot p(y_2) \cdot ... p(y_N) = L(y_1, y_2, ..., y_N)$$

- E' il prodotto delle probabilità semplici.

- Questa è la **Funzione di verosimiglianza** o **funzione di Likelihood**, L(.)

Funzione di verosimiglianza (riassunto)

- Data una serie di misure y_i i=1…N di variabili casuali…
- … Note le densità di probabilità di ciascuna variabile casuale…
- … Sotto l'ipotesi che le variabili siano tra loro indipendenti…
- … E' possibile scrivere la funzione di verosimiglianza come il prodotto delle probabilità di ciascuna misura y_i i=1…N.

Stima alla massima verosimiglianza caso Gaussiano

- Supponiamo il vettore **y** corrisponda a N realizzazioni di una variabile gaussiana a media μ, deviazione standard σ (N misure indipendenti di una stessa quantità)
- La funzione verosimiglianza dipende da μ e σ.

$$L(y_1, y_2,..., y_N \mid \mu,\sigma) = p(y_1)\cdot p(y_2)\cdot ...p(y_N) =$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}}\cdot \exp\left[-\frac{1}{2}\left(\frac{y_1-\mu}{\sigma}\right)^2\right]\cdot ...\frac{1}{\sqrt{2\pi\sigma^2}}\cdot \exp\left[-\frac{1}{2}\left(\frac{y_N-\mu}{\sigma}\right)^2\right] =$$

Stima alla massima verosimiglianza

- Se **massimizziamo** L=L(**y** | μ,σ) rispetto a μ e σ
- troviamo i parametri μ,σ tali per cui è massima la probabilità di misurare il vettore di dati **y = {**y_i, i=1…N}.
- **Stima alla massima verosimiglianza**.
- Più in generale, le variabili possono avere densità di probabilità diverse, ciascuna descritta da un set di parametri stimabili con l'approccio alla massima verosimiglianza…
- La funzione di verosimiglianza dipende dai parametri che definiscono le densità di probabilità delle variabili casuali che entrano nella verosimiglianza…
- Massimizzando la funzione di verosimiglianza rispetto a tali parametri se ne effettua la stima in modo tale che il vettore osservato yi i=1…N sia massimamente probabile (massima verosimiglianza).

Stima alla massima verosimiglianza
Il caso gaussiano

- E' solitamente più facile minimizzare il logaritmo negativo della verosimiglianza, f(.), (prodotto → sommatoria)

$$f(y_1, y_2, ..., y_N \mid \mu, \sigma) = -\ln[L(y_1, y_2, ..., y_N \mid \mu, \sigma)] =$$

$$= -\ln\left\{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right]\right\} =$$

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

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

Determino μ

- Per trovare il minimo, poniamo a zero le derivate:

$$\frac{\partial f(y_1, y_2 .... y_N \mid \mu, \sigma)}{\partial \mu} = \frac{\partial}{\partial \mu}\left\{N \cdot \ln\left(\sqrt{2\pi}\right) + N \cdot \ln(\sigma) + \sum_{i=1}^{N}\left[\frac{1}{2} \cdot \left(\frac{y_i-\mu}{\sigma}\right)^2\right]\right\} =$$

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

$$\sum_{i=1}^{N}(y_i - \mu) = 0 \Rightarrow \sum_{i=1}^{N} y_i = N \cdot \mu \Rightarrow \mu = \frac{\sum_{i=1}^{N} y_i}{N}$$

Media campionaria!

Determino σ

- Per trovare il minimo, poniamo a zero le derivate:

$$\frac{\partial f(y_1, y_2, ..., y_N \mid \mu, \sigma)}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left\{ N \cdot \ln\left(\sqrt{2\pi}\right) + N \cdot \ln(\sigma) + \sum_{i=1}^{N}\left[\frac{1}{2}\cdot\left(\frac{y_i - \mu}{\sigma}\right)^2\right]\right\} =$$

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

$$= \frac{N}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{N}(y_i - \mu)^2 = 0 \Rightarrow N - \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \mu)^2 = 0 \Rightarrow$$

$$N \cdot \sigma^2 = \sum_{i=1}^{N}(y_i - \mu)^2 \Rightarrow \sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}$$

Varianza campionaria!

Overview

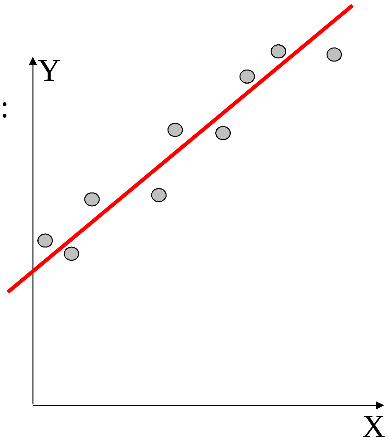
Distribuzioni di probabilità

Stima alla massima verosimiglianza

Stima dei parametri di una retta

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 ai minimi quadrati e verosimiglianza

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

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

Stima ai minimi quadrati e verosimiglianza

Scriviamo il logaritmo negativo della verosimiglianza:

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

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

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

Stima ai minimi quadrati e verosimiglianza

- E massimizziamolo ponendo a zero le derivate rispetto a m e b:

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

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

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

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

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

1ª equazione

Stima ai minimi quadrati e verosimiglianza

E massimizziamolo ponendo a zero le derivate rispetto a m e b:

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

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

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

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

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

Stima massima verosimiglianza

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

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

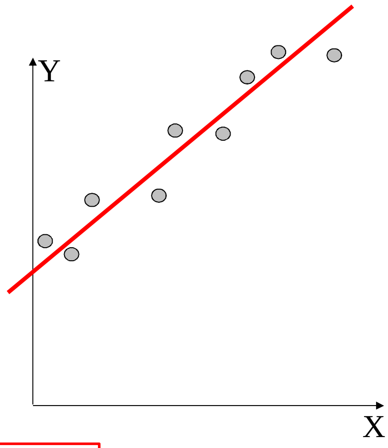
Le incognite, m e b, compaiono con esponente 1 => equazioni lineari in m e b
Potrei risolvere per sostituzione

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?

Overview

Distribuzioni di probabilità

Stima alla massima verosimiglianza

Sistemi lineari

Sommario

Matrici e Sistemi lineari

Esempio di sistema linearizzato

Soluzione di un sistema lineare

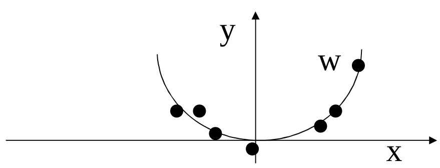
Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare

Modello

$y=f(x; w)$

$y = a\ x^2$



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

Controllo: Utilizzo il modello (w noti) per predire l'uscita, y, in funzione dell'ingresso x.

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** e **y** vettori.

Sistema lineare

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

....

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

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

....

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

Matrici

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

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

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

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)

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

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 $\neq$ BA

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

$$AI = A = IA$$

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

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

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(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 $U^T U = diag(W)$.
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)**.

Sommario

Matrici e Sistemi lineari

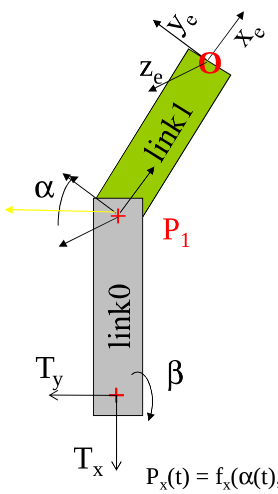
Esempio di sistema linearizzato

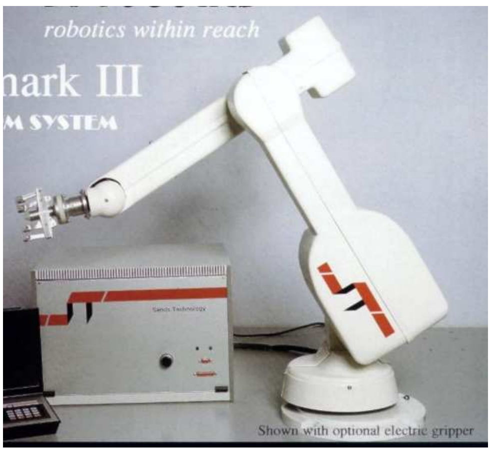
Soluzione di un sistema lineare

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare

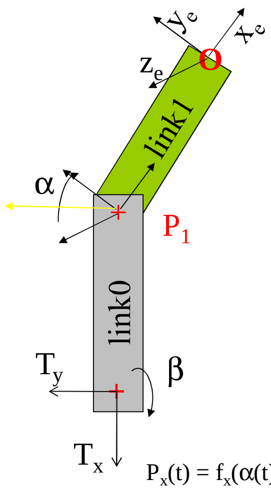
Esempio di "sistema"

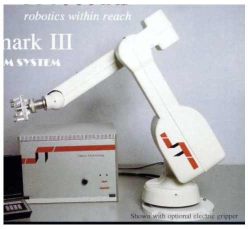




$P_x(t) = f_x(\alpha(t), \beta(t), T_x(t), T_y(t) | l_0, l_1).$
$P_y(t) = f_y(\alpha(t), \beta(t), T_x(t), T_y(t) | l_0, l_1).$
$P_z(t) = f_z(\alpha(t), \beta(t), T_x(t), T_y(t) | l_0, l_1).$

Esempio di “sistema"



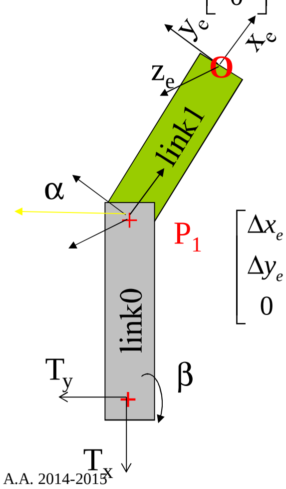


Le funzioni legano la posizione dell’end point, uscita **P**, alla posizione degli angoli, α e β e della posizione della base, **T**, che rappresentano gli ingressi.

$P_x(t) = f_x(\alpha(t), \beta(t), T_x(t), T_y(t) | l_0, l_1).$
$P_y(t) = f_y(\alpha(t), \beta(t), T_x(t), T_y(t) | l_0, l_1).$
$P_z(t) = f_z(\alpha(t), \beta(t), T_x(t), T_y(t) | l_0, l_1).$

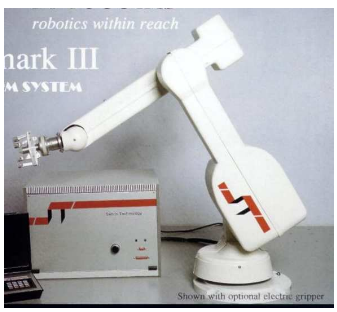
Rappresentazione linearizzata
Sistema lineare

$$\begin{bmatrix} \Delta x_e \\ \Delta y_e \\ 0 \end{bmatrix} = \begin{bmatrix} -l_1 \sin(\alpha+\beta) & -l_1 \sin(\alpha+\beta) - l_0 \sin\beta & 1 & 0 \\ -l_1 \cos(\alpha+\beta) & -l_1 \cos(\alpha+\beta) - l_0 \cos\beta & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta\alpha \\ \Delta\beta \\ \Delta T_x \\ \Delta T_y \end{bmatrix}$$



$\alpha = 90$ $\quad l_0 = 2{,}5$
$\beta = 0$ $\quad l_1 = 2$

$$\begin{bmatrix} \Delta x_e \\ \Delta y_e \\ 0 \end{bmatrix} = \begin{bmatrix} -2 & -2 & 1 & 0 \\ 0 & -2.5 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta\alpha \\ \Delta\beta \\ \Delta T_x \\ \Delta T_y \end{bmatrix}$$



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

Sommario

Matrici e Sistemi lineari

Esempio di sistema linearizzato

Soluzione di un sistema lineare

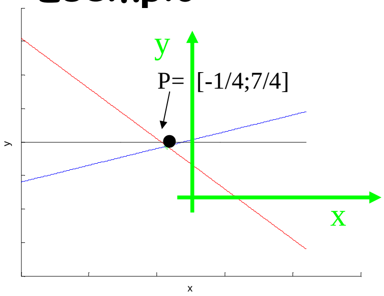
Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare

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: $x_1 = -2 + x_2$.

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

$x_1 - 1/4 = 2 \rightarrow x_1 = -1/4$

Sistema lineare

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

....

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

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

....

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

A x = b

M x 1

Vettore dei termini noti

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

Sistema <u>quadrato</u> (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

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

Scrivo il sistema lineare: $Ax = b$

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

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

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

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 $\det(A) \neq 0$

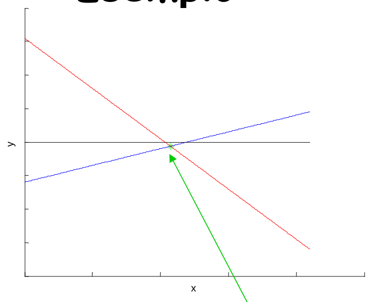
Il sistema ammette 1 ed 1 sola soluzione se $\det(A) \neq 0$

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

$a_{11}x_1+a_{12}x_2 = b_1$

$a_{21}x_1+a_{22}x_2 = b_2$

$\mathbf{y = Ax}$

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

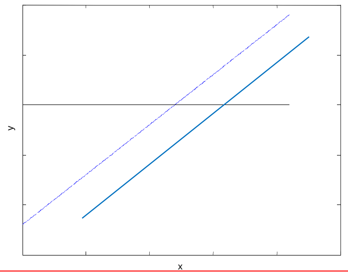
$$A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

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

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

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

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

....

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

Esempio:

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

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

....

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

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

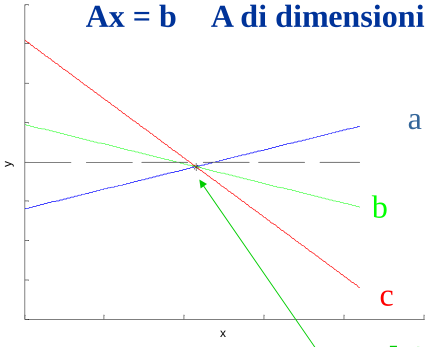
J(W,L) è rettangolare: numero di righe maggiore del numero di colonne

$y = x + 2$

$y = -3x + 1$

$y = -x + 3/2$

Ax = b   A di dimensioni m x n



P = [-1/4 7/4]

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

Rango di A è pieno

Rango di una matrice

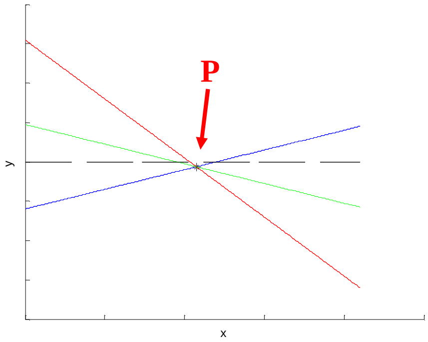
det (A§ij)   Minore complementare

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)



α_1 (y - x - 2) +
α_2 (y +3x – 1) =
(y + x - 3/2)

In questo caso:
α_1 = -1/2
α_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:

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

$$\Downarrow$$

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

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

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

$y = x - 2$

$y = -3x + 1$

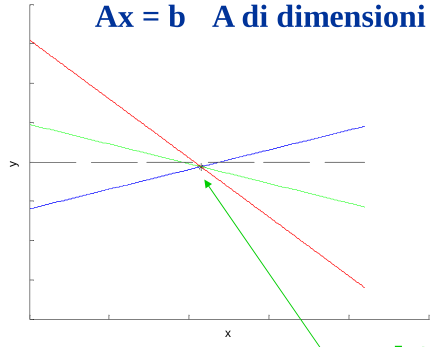
$y = -x + 3/2$

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

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

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

Ax = b A di dimensioni m x n



P = [-1/4 7/4]

$P = C * A^T * b$ **P = [-0.25 +1.75]**

intersezione

Riformulazione del problema con rumore

$a_{11}x_1 + a_{12}\,x_2 + \ldots\ldots\ a_{1N}\,x_N = b_1 + v_1$

$a_{21}x_1 + a_{22}\,x_2 + \ldots\ldots\ a_{2N}\,x_N = b_2 + v_2$

....

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

Modello **Misure**

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

A x = b + 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: $(Ax - b)^2 = \sum_k v_k^{\,2} = \|Ax - b\|^2$

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.

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

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

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

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

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.

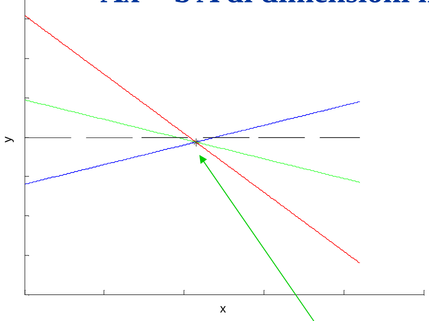
Sistemi lineari con m > n

y = x - 2

y = -3x + 1

y = -x + 3/2

Ax = b A di dimensioni m x n



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

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

P = [-1/4 7/4]

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

$P = C * A^T * b$ **P = [-0.25 +1.75]**

intersezione

$\|Ax - b\| = 0$

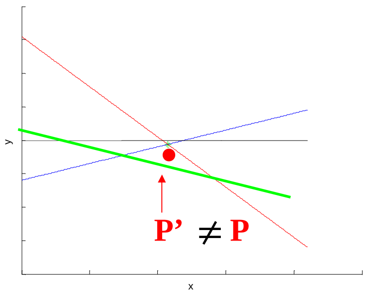
Sistemi lineari con m > n - non esiste soluzione (matematica)

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

y = x + 2

y = -3x + 1

y = -x + 1/2



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

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

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

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

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

P' = [-0.5 +1.4167]

No intersezione

Commenti

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

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

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

Lo scarto misura la distanza (verticale) dalla retta

Sommario

Matrici e Sistemi lineari

Esempio di sistema linearizzato

Soluzione di un sistema lineare

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare