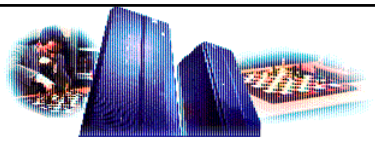
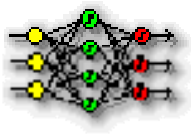




Sistemi Intelligenti
I sistemi lineari, Tecniche di base per l'ottimizzazione non-lineare

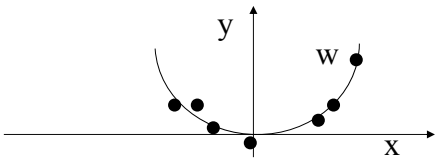
Alberto Borghese
Università degli Studi di Milano
Laboratorio di Sistemi Intelligenti Applicati (AIS-Lab)
Dipartimento di Informatica
borghese@di.unimi.it

Sommario

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

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

.....

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

{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:
$3x_1 + 2\ x_2 + \ldots\ldots\ 4x_N = 5$
$4x_1 - 2\ x_2 + \ldots\ldots\ 0.5\ x_N = 3$

.....

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

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)

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

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 ≠ BA

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

$$AI = A = IA$$

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

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

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 $U^T U = I$ ➔ $U^{-1} = U^T$

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

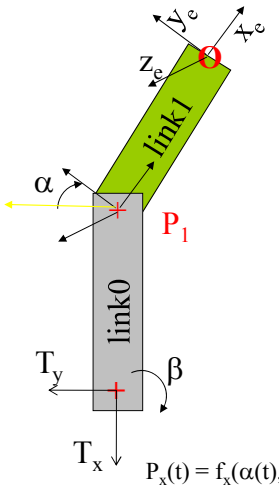
Esempio di sistema linearizzato

Soluzione di un sistema lineare

Analisi dell'affidabilità della stima

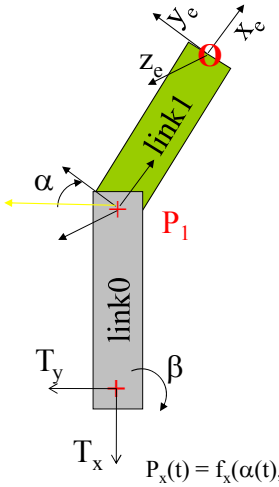
Determinazione dei parametri di un modello non-lineare

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"

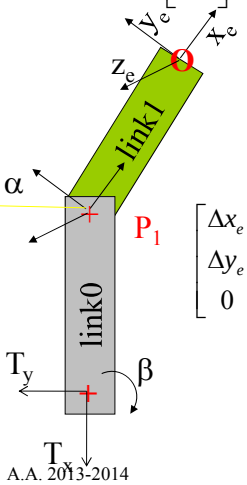


Le funzioni legano la posizione dell'end point, uscita **P**, alla posizione degli angoli, α e β e della posizione della base, **T**, che rappresentano gli ingressi.

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

Rappresentazione linearizzata Sistema lineare

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



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

$\beta = 0$ $\quad l_1 = 2$

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

b = A x

Sommario

Matrici e Sistemi lineari

Esempio di sistema linearizzato

Soluzione di un sistema lineare

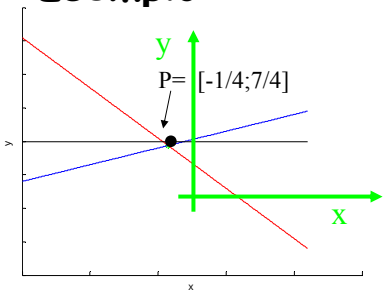
Analisi dell'affidabilità della stima

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: $x_1 = -2 + x_2$.

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

$x_1 - 1/4 = 2$ ➔ $x_1 = -1/4$

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:

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

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

.....

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

A x = b

M x 1

Vettore dei termini noti

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

Sistema quadrato (N x N)

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

.....

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

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

.....

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

Ammette 1, nessuna o ∞ soluzioni

A è N x N quadrata

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

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

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

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

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

Soluzioni:
- ! $\exists$ Soluzione (sistema impossibile)
- $\exists$ 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

$y = x + 2$

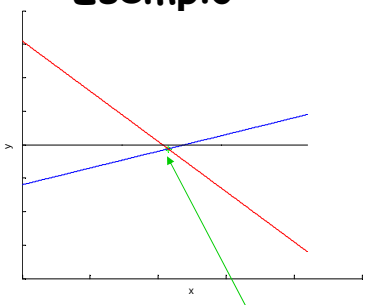
$y = -3x + 1$

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

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

$x_1 = x$

$x_2 = y$



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

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

Rango di A è pieno

$x_1 = -1/4$

$x_2 = 7/4$

$P = A^{-1} b$

P = [-1/4 7/4]

Risoluzione di un sistema 2x2

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

$\mathbf{y} = \mathbf{A}\mathbf{x}$

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

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

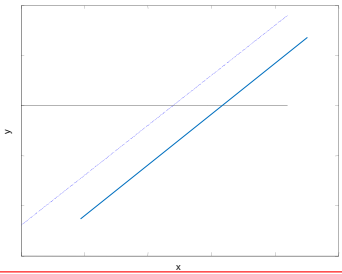
$\det(A) = a_{11}*a_{22}-a_{12}*a_{21}$

Esempio di soluzione non univoca

$y = x + 2$
$2y = 2x + 3$

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

Non esistono soluzioni



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

$x_1 = x$
$x_2 = y$

$\det(A) = 1(-2)-(-1)(2)=-2+2=0$ 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:
$3x_1 + 2x_2 + \ldots\ldots 4x_N = 5$
$4x_1 - 2x_2 + \ldots\ldots 0.5x_N = 3$

……

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

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.

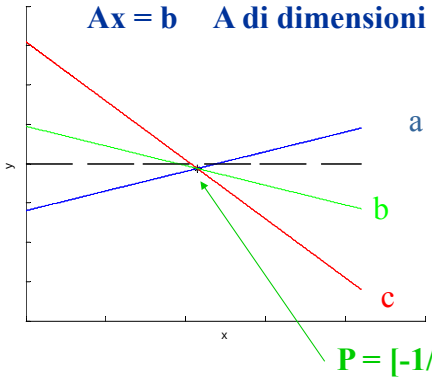
Sistemi lineari con m > n

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

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

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

Ax = b A di dimensioni m x n



P = [-1/4 7/4]

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

Esiste un'equazione "di troppo"

Nessuna, 1 o ∞ soluzioni

Rango di A è pieno

Rango di una matrice

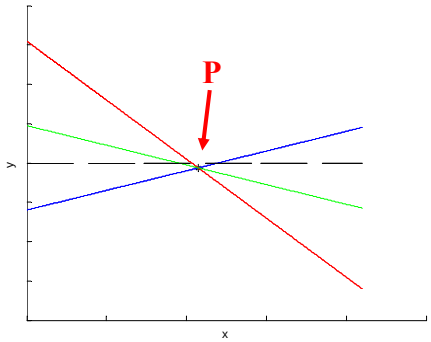
det ($A^{\S}ij$) Minore complementare

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)



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:

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

$\Downarrow$

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

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

Quale criterio viene soddisfatto da **x**?

Sistemi lineari con m > n

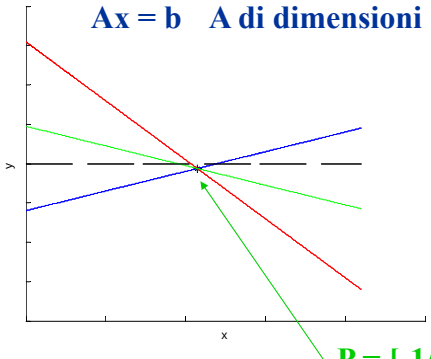
$y = x - 2$

$y = -3x + 1$

$y = -x + 3/2$

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

Ax = b A di dimensioni m x n



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

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

Riformulazione del problema con rumore

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

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

.....

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

Modello **Misure**

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

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

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

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.

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

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

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

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

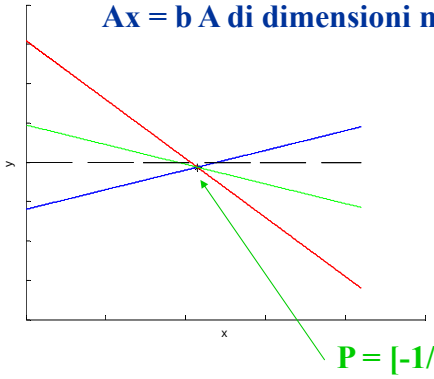
NB le funzioni costo sono spesso quadratiche (problemi di minimizzazione convessi) perchè il costo cresce sia che il modello sovrastimi che sottostimi le misure.
Inoltre, le derivate calcolate per imporre le condizioni di stazionarietà (minimo), sono relativamente semplici.

Sistemi lineari con m > n

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

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

Ax = b A di dimensioni m x n



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

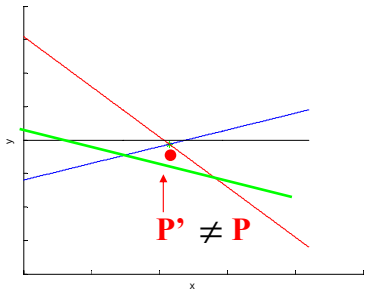
$||Ax - b|| = 0$

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

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

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

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



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

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

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

$P = C * A^T * b$ **P' = [-0.5 +1.4167]**

No intersezione

Commenti

$$\sum_k \nu_k^{\,2} = \|Ax-b\|^2 = \sum_k \| A_{k,*}x - b_k \|^2 =$$
$$\left[(A_{11}x_1 + A_{12}x_2) - b_1\right]^2 + \left[(A_{21}x_1 + A_{22}x_2) - b_2\right]^2 +$$
$$\left[(A_{31}x_1 + A_{32}x_2) - b_3\right]^2$$

Lo scarto misura la distanza (verticale) dalla retta

Stima ai minimi quadrati pesata

$\min \| P(Ax - b) \|$ **P di dimensioni m x m – matrice dei pesi, diagonale**

$p_1a_{11}x_1 + p_1a_{12}x_2 - p_1b_1 = p_1\nu_1$
$p_2a_{21}x_1 + p_2a_{22}x_2 - p_2b_2 = p_2\nu_2$
$p_3a_{31}x_1 + p_3a_{32}x_2 - p_3b_3 = p_3\nu_3$

Residuo pesato $\min \sum_k (p_k\nu_k)^2$

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

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

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

Condizionamento della matrice C = A'*A

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

Per evitare di ottenere elementi troppo grandi che rendono la norma della matrice C vicina alla precisione della macchina, si preferisce utilizzare la Singular Value Decomposition per risolvere il sistema lineare.

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

Sistema lineare: soluzione robusta

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

Numero di condizionamento varia circa con (A'*A).

Soluzione tramite Singular Value Decomposition (diagonalizzazione)

Numero di condizionamento varia circa con det(A).

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

$\mathbf{U\,W\,V\,X = B}$

$$\mathbf{x = V'W^{-1}U'b}$$

Ortonormale M x N — Diagonale (N x N) — Ortonormale N x N

$\mathbf{V^T\,W^{-1}\,U^T\,U\,W\,V\,X = V^T\,W^{-1}\,U^T\,b} \Rightarrow \mathbf{X = V^T\,W^{-1}\,U^T\,b}$

- La matrice C non viene formata.
- W^{-1} contiene i reciproci degli elementi di W.

W^{-1} è diagonale. $w_{ii}^{-1} = 1/w_{ii}$

Rank-deficiency nella matrice dei coefficienti

Quando C è singolare?

$x = (A'^*A)^{-1}A'^* b$ $\quad$ **$x = V'W^{-1}U'b$**

Se A è rank-deficient, A'*A è singolare.

Si può facilmente osservare valutando il valore singolare più piccolo della matrice W che risulta uguale a 0.

In questo caso il problema è sovraparametrizzato.

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

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**.

Valutazione della bontà della stima

$$x = (A'^*A)^{-1}A'^* b \quad <==> \quad \min_{\mathbf{x}} \sum_k \nu_k^2 = \min_{\mathbf{x}} (\mathbf{Ax}-\mathbf{b})^2$$

Errore di modellizzazione Gaussiano a media nulla $N(0,\sigma^2)$

$$<v_k> = 0$$

$$\hat{\sigma}_0^2 = \sum_{k=1}^{M} \left(\nu_k^2\right) = |v|^2$$

Varianza della stima = varianza dell'errore di misura

.

Valutazione della bontà della stima del singolo parametro e della loro correlazione

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

$x = CA'^* b$

$$\hat{\sigma}_0^2 = \sum_{m=1}^{M} \left(v_m^2\right)$$

Chiamiamo u e v le variabili casuali associate all'errore sui parametri e all'errore di modellizzazione, rispettivamente. Si suppone errore a media nulla e Gaussianamente distribuito.

$u = \Delta x$ $\quad (x + u) = C A' (b + v)$

$\Downarrow$

$\mathbf{x = CA'b}$ $\quad u = C A'^* v$ $\quad E[u] = 0$

Impostazione del calcolo della correlazione tra i parametri

$u = C A' v$

Vogliamo individuare la correlazione tra due parametri i e j. Devo quindi determinare la loro correlazione:

$$\begin{bmatrix} u_1^2 & u_1u_2 & \dots & u_1u_W \\ u_2u_1 & u_2^2 & \dots & u_2u_W \\ \dots & \dots & \dots & \dots \\ u_Wu_1 & u_Wu_2 & \dots & u_W^2 \end{bmatrix}$$

$\langle u_i, u_j \rangle$

$u = C A' v \quad \Rightarrow \quad u' = v'A (C)'$

$uu' = C A' vv'A C'$ => Applicando l'operatore di media, si ottiene:

$\langle uu' \rangle = C A' \langle vv' \rangle A C'$

Dato che v sono i residui, e sono indipendenti, e tutte i punti di controllo hanno lo stesso tipo di errore di misura, si avrà che $\langle vv' \rangle = I\sigma_0^2$.

Correlazione tra i parametri

$\langle uu' \rangle = CA'\, IA\, C' \sigma_0^2 = C'\, \sigma_0^2$ $\qquad$ $\boxed{\langle u'u \rangle = C\, \sigma_0^2}$

Da cui si giustifica il nome di matrice di covarianza per C.

Segue che: $\sigma^2(u_{ij}) = c_{ij}\, \sigma_0^2$ Varianza sulla stima del parametro.

$$-1 \le r_{ij} = \frac{< u_i u_j >}{\sqrt{< u_i >^2 < u_j >^2}} = \frac{c_{ij}}{\sqrt{c_i c_j}} \le +1$$

Indice di correlazione tra il parametro i ed il parametro j
(empiricamente si scartano parametri quando la correlazione è superiore al 95%)

Vanno rapportati alle dimensioni dei parametri coinvolti.

Matrice di covarianza

Date N variabili casuali: $x = [x_1, x_2, \ldots x_N]$ si può misurare la correlazione tra coppie di variabili. E' comodo rappresentare la correlazione tra variabili casuali in un'unica matrice detta **matrice di covarianza** come:

$$C = \begin{bmatrix} \sigma_{x_1x_1} & \sigma_{x_1x_2} & . & \sigma_{x_1x_N} \\ \sigma_{x_2x_1} & \sigma_{x_2x_2} & . & \sigma_{x_2x_N} \\ . & . & . & . \\ \sigma_{x_Nx_1} & \sigma_{x_Nx_2} & . & \sigma_{x_Nx_N} \end{bmatrix}$$

Varianza: $\sigma_{x_1x_1} = \sigma^2_{x_1}$ N parametri

Covarianza: $\sigma_{x_ix_j} = \sigma_{x_jx_i}$ $i \neq j$ $(N-1)^2/2$ parametri

Correlazione

Date due variabili casuali: x_i, x_j, l'indice di correlazione misura quanto le coppie di variabili estratte: $p(x_i, x_j)$ stanno su una retta:

$$r = \frac{M_{x_i x_j} - M_{x_i} M_{x_j}}{\sigma_{x_i} \sigma_{x_j}} \qquad -1 \le r \le +1$$

Definendo la covarianza tra x_i ed x_j come:

$$\sigma_{x_i x_j} = \frac{1}{N} \sum_i \sum_j \left(x_i - M_{x_i}\right)\left(x_j - M_{x_j}\right)$$

Dalla definizione di deviazione standard risulta:

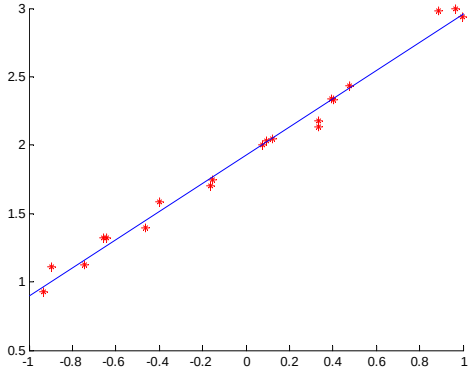
$$r = \frac{\sigma_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}}$$

Caso 2D

N = 20 punti $\sigma_o^2 = 0.01$
m reale = 1 q reale = 2

C =
0.1427 -0.0002
-0.0002 0.0500

m stimato = 1.0302
q stimato = 1.9308



C=
0.1702 0.0124
0.0124 0.0509

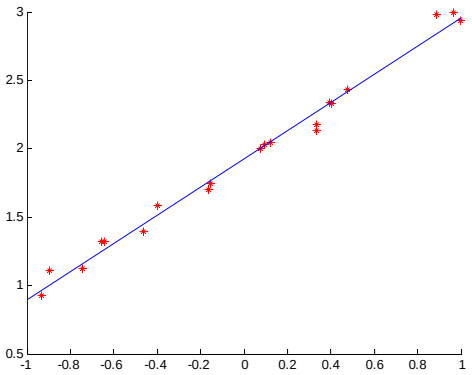
m stimato = 0.9937
q stimato = 1.9522

$y = mx + q$

Caso 2D – less points

N = 10 punti $\sigma_o^2 = 0.01$
m reale = 1 q reale = 2

C =
0.5927 -0.0030
-0.0030 0.1000
m_stimato =
1.0081
q_stimato =
1.9616



C =
0.2514 -0.0360
-0.0360 0.1051
m_stimato =
1.0012
q_stimato =
1.9107

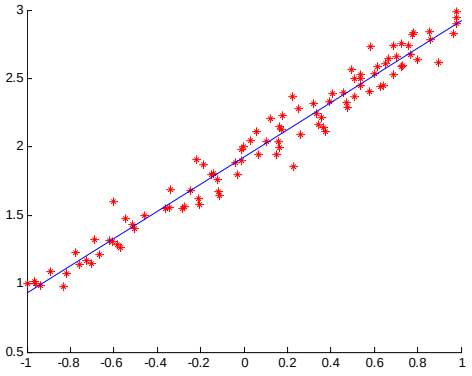
$y = mx + q$

Diminuisce la confidenza nella stima

Caso 2D – more points

N = 100 punti $\sigma_o^2 = 0.01$
m reale = 1 q reale = 2

C =
0.0327 -0.0034
-0.0034 0.0103
m_stimato =
0.9935
q_stimato =
1.9270



C =
0.0310 0.0023
0.0023 0.0102
m_stimato =
0.9776
q_stimato =
1.9285

$y = mx + q$

Aumenta la confidenza nella stima

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

Stima di parametri in insiemi di equazioni non lineari - linearizzazione

y = f(x) viene linearizzata utilizzando il differenziale:

$$y = f(x_o) + \left.\frac{df(x)}{dx}\right|_{x=x_o} dx = y_o + \left.\frac{df(x)}{dx}\right|_{x=x_o} dx$$

Si può vedere come sviluppo di Taylor arrestato al 1° ordine
E' un'equazione lineare in dx.

Per funzioni di più variabili, f(**P;W**) = 0, la linearizzazione si può scrivere come:

$$F(\mathbf{P};\mathbf{W}) = F(\mathbf{P_o};\mathbf{W_o}) + \sum_{j=1}^{W} \left.\frac{\partial F(.)}{\partial w_j}\right|_{P_o,W_o} * dw_j = \quad k - \sum_{j=1}^{W} a_j * dw_j$$

E' un'equazione lineare nei dw che descrive il comportamento della funzione F(.) nell'intorno del punto Po con i parametri Wo.

Metodo di Gauss-Newton

- L'idea:

Inizializzazione:

- Inizializzo i parametri ad un valore iniziale.

Iterazioni:

- 1) Linearizzazione delle equazioni.
- 2) Stima dell'aggiornamento dei parametri nel modello linearizzato ai minimi quadrati (soluzione ottimale, minimo del problema linearizzato).
- 3) Correzione dei parametri.

Può essere pesante perchè richiede l'inversione della matrice di covarianza.
Spesso si preferiscono utilizzare metodi di ottimizzazione del primo ordine.

In pratica

$\mathbf{y} = f(\mathbf{x})$ x, y vettori di N ed M elementi rispettivamente

$\mathbf{y_0} = f(\mathbf{x_0})$ x_0, y_0 valore iniziale

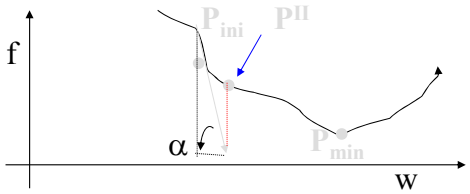
Iterazione di (nella prima iterazione k = 0):

- $\mathbf{dy_k} + \mathbf{y_k} = (\Sigma \delta f(\mathbf{x})/d\mathbf{x})_{\mathbf{xk}}\ \mathbf{dx} + \mathbf{f(x_k)}$ $(\Sigma \delta f(\mathbf{x})/d\mathbf{x})_{\mathbf{xk}}$ **are numbers!**
- Si ottiene un sistema lineare
- Viene risolto come $\mathbf{dx_k} = (AA^T)^{-1}A^T\mathbf{dy_k}$
- **Si aggiorna il valore di x come $\mathbf{x_{k+1}} = \mathbf{x_k} + \mathbf{dx_k}$**

Fino a convergenza

Minimizzazione tramite gradiente (metodo del primo ordine): 1 variabile

Tecnica del gradiente applicata alla minimizzazione di funzioni non-lineari di **una variabile, x,** e di **un parametro, w**: f = f (**x** | **w**).



La derivata, mi dà due informazioni:

1) In quale direzione di **w**, la funzione decresce.
2) Quanto rapidamente decresce.

Definisco uno spostamento arbitrario lungo la pendenza: maggiore la pendenza maggiore lo spostamento.

dw ∝ -f'(w;P) dati P, w. La derivata viene calcolata rispetto a w.

Occorre un'inizializzazione.

Metodo iterativo.

mi.it\

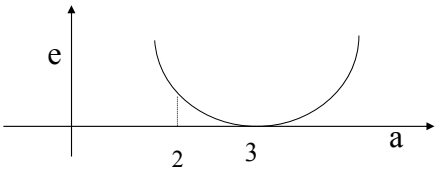
Esempio di applicazione tecnica del gradiente per funzioni di 1 variabile

Supponiamo che il modello da noi considerato sia semplice: $y = ax^2$

Abbiamo un unico parametro da determinare: a. La funzione è lineare in a.

Misuriamo un punto sulla parabola: x = 1, y =3.
Vogliamo modificare a in modo che la parabola passi per P(x,y).
La funzione costo da minimizzare sarà: $e = f(a \mid x,y) = (y-ax^2)^2$
La soluzione è a = 3

Partiamo da $a_{ini} = 2$.
$err = (3 - 2*1)*1 = 1$



Utilizziamo il metodo del gradiente:
Calcoliamo la derivata di $f(a \mid x,y)$ ➔ $f'(a) = -2 (y - a x^2) x^2$

Minimizzazione - underdamping

Consideriamo $\alpha = 1$
Calcoliamo la derivata di f(.) ➔ $f'(.) = -2 (y - a x^2) x^2$

Utilizziamo il metodo del gradiente:

Passo 1:
Calcoliamo l'incremento da dare al parametro a:
da =-[-2 (3 – 2 1) 1] = -[-6 +4] = 2 a' = 2 + 2 = 4

Passo 2:
Calcoliamo l'incremento da dare al parametro a:
da = -[-2 (3 - 4 1) 1] = -[-6 +8] = -2 a'' = 4 – 2 = 2
Oscillazioni!!!

Mi sposto troppo velocemente da una parte all'altra del minimo.

Minimizzazione -2 passi

Consideriamo $\alpha = 0.4$
Calcoliamo la derivata di f(.) ➔ $f'(.) = -2 (y - a x^2) x^2$

Utilizziamo il metodo del gradiente:

Passo 1:
Calcoliamo l'incremento da dare al parametro a:
da = -0.4 [-2 (3 – 2 1) 1] = -[-6 +4] = 0.8 a' = 2 + 0.8 = 2.8

Passo 2:
Calcoliamo l'incremento da dare al parametro a:
da = - 0.4 [-2 (3 – 2.8 1) 1] = -[-6 + 5.6] = 0.16 a'' = 2.8 + 0.16 = 2.96
Converge ad a = 3.

Posso correre il rischio di spostarmi troppo lentamente

Minimizzazione di funzioni di più variabili

min(f(**x**, **w**)) funzione costo od errore, **w** vettore.

Modifico il valore dei pesi di una quantità proporzionale alla pendenza della funzione costo rispetto a quel parametro.
La pendenza è una direzione nello spazio, non è più solamente destra / sinistra. Devo calcolare la derivata spaziale = **gradiente** della funzione costo, f(.).
Estensione della tecnica del gradiente a più variabili.

$dw = - \alpha \nabla f(x;w)$, dato P, W.

Serve un'**approssimazione iniziale** per i pesi $W_{ini} = \{w_j\}_{ini}$.

Evoluzione dei metodi del primo ordine

- α è un parametro critico. Se è troppo piccolo convergenza molto lenta, se è troppo grande overshooting.
- Ottimizzazione di α. Ad ogni passo viene calcolato a ottimale, per cui la funzione è decrescente (line search).

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