

Sistemi Intelligenti I sistemi lineari

Alberto Borghese

Università degli Studi di Milano
Laboratorio di Sistemi Intelligenti Applicati (AIS-Lab)
Dipartimento di Scienze dell'Informazione
borghese@dsi.unimi.it



A.A. 2008-2009

1/59

<http://homes.dsi.unimi.it/~borghese/>



Sommario



Matrici

Sistemi lineari

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare

A.A. 2008-2009

2/59

<http://homes.dsi.unimi.it/~borghese/>



Sistema lineare



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

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

.....

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

{ a_{ij} } – coefficienti in numero $N \times M$

{ x_j } – incognite, M

{ b_j } – termini noti, N

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

Esempio:

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

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

.....

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



Matrici



$$A = [a_{i,j}]$$

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

$$\alpha A = [\alpha a_{i,j}]$$

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

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

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

Se $A (n \times m) \rightarrow B (m \times p) \rightarrow C (n \times p)$

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

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:

$$(A+B)C = AC + BC.$$

$$AB \neq BA$$

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

$$AI = A = IA$$

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

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



Minore complementare



$$A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

A_{ij}^* minore complementare di a_{ij} = determinante della matrice ottenuta eliminando la riga i e la colonna j di A .

$$A_{21}^* = \det \begin{bmatrix} 3 & -2 \\ 1 & 2 \end{bmatrix} = 3*2 - (-2*1) = +8$$



Determinante di una matrice Quadrata



$$\det(A) = \sum_i (-1)^{(i+j)} a_{ij} A^*_{ij} = \sum_j (-1)^{(i+j)} a_{ij} A^*_{ij}$$

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

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \longleftarrow \text{Elementi sulla riga}$$

$$\det(A) = (-1)^{(2+1)} (2) [(3 * 2) - (-2 * 1)] + (-1)^{(2+2)} (0) [(1*2) - (-2*1)] + (-1)^{(2+3)} (1) [(1*1) - (3*1)] = -16 + 2 = -14$$



Calcolo della matrice inversa



Matrice dei complementi algebrici

$$A^{-1} = [1/\det(A)] A^+$$

$$A^+_{ij} = (-1)^{i+j} \det(A^*_{ji})$$

Minore complementare della matrice A trasposta



Esempio di matrice Inversa



$A = [a_{ij}]$, matrice quadrata.

A^+_{ij} matrice dei
complementi algebrici =
minori complementari
moltiplicati $(-1)^{i+j}$

$$A^{-1} = 1/\det(A) \begin{bmatrix} A^+_{11} & A^+_{21} & A^+_{n1} \\ A^+_{12} & A^+_{22} & A^+_{32} \\ A^+_{13} & A^+_{23} & A^+_{33} \end{bmatrix} \quad A^{-1} A = I$$

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad A^{-1} = \frac{-1}{14} \begin{bmatrix} 0 \cdot 2 - 1 \cdot 1 & -[3 \cdot 2 - (-2) \cdot 1] & 3 \cdot 1 - (-2) \cdot 0 \\ -[2 \cdot 2 - 1 \cdot 1] & 1 \cdot 2 - (-2) \cdot 1 & -[1 \cdot 1 - (-2) \cdot 2] \\ 2 \cdot 1 - 0 \cdot 1 & -[1 \cdot 1 - 3 \cdot 1] & 1 \cdot 0 - 3 \cdot 2 \end{bmatrix} = (-1/14) \begin{bmatrix} -1 & -8 & 3 \\ -3 & 4 & -5 \\ 2 & 2 & -6 \end{bmatrix}$$

$$\det A = -14 \quad AA^{-1} = -1/14 \begin{bmatrix} 1(-1) + 3(-3) - 2(-6) & 1(-8) + 3 \cdot 4 - 2(-2) & 1 \cdot 3 + 3(-5) - 2(-6) \\ 2(-1) + 0(-3) + 1(2) & 2(-8) + 0 \cdot 4 + 1 \cdot 2 & 2 \cdot 3 + 0(-5) + 1(-6) \\ 1(-1) + 1(-3) + 2 \cdot 2 & 1(-8) + 1 \cdot 4 + 2 \cdot 2 & 1 \cdot 3 + 1(-5) + 2(-6) \end{bmatrix} = I$$

Se esiste, la matrice inversa è unica.



Altre proprietà delle matrici



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

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

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

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

Una matrice U , si dice ortogonale se $U^T U = \text{diag}(W)$.

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).**



Rango di una matrice

Data una matrice A di ordine n ($n \times n$),

una matrice A $n \times n$ 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 $n \times n$ ha rango n (rango pieno) se e solo se
il suo determinante è diverso da 0

Rango di una matrice $M \times N$ 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.



Sommario

Matrici

Sistemi lineari

Analisi dell'affidabilità della soluzione

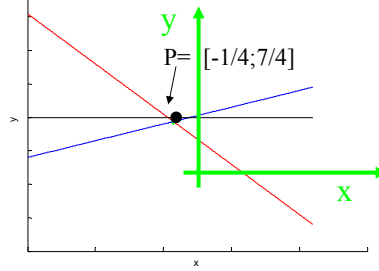
Determinazione dei parametri di un modello non-lineare



Esempio

$$y = x + 2$$

$$y = -3x + 1$$



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

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

$$y = x_2$$

$$x = x_1$$

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

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

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

A.A. 2008-2009

13/59

<http://homes.dsi.unimi.it/~borghese/>



Sistema lineare

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

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

.....

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

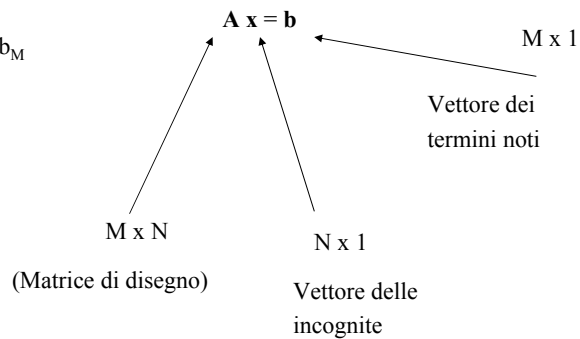
Esempio:

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

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

.....

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



A.A. 2008-2009

14/59

<http://homes.dsi.unimi.it/~borghese/>



Sistema quadrato (N x N)



$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1N}x_N &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2N}x_N &= b_2 \end{aligned}$$

Ammette 1, nessuna o ∞ soluzioni

A è N x N quadrata

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

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

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

altrimenti, **nessuna** (rette parallele)

o

∞ **soluzioni** (rette coincidenti).

.....

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

Esempio:

$$\begin{aligned} 3x_1 + 2x_2 + \dots + 4x_N &= 5 \\ 4x_1 - 2x_2 + \dots + 0.5x_N &= 3 \end{aligned}$$

.....

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



Soluzione dei sistemi lineari



Scrivo il sistema lineare: $\mathbf{Ax} = \mathbf{b}$

$$y = x + 2$$

$$y = -3x + 1$$

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

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

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

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

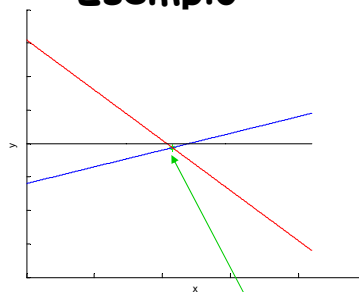


Esempio



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

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



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

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

$$x_1 = x$$

$$x_2 = y$$

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

Rango di A è pieno

$$x_1 = -1/4$$

$$x_2 = 7/4$$

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

$$P = \begin{bmatrix} -1/4 & 7/4 \end{bmatrix}$$

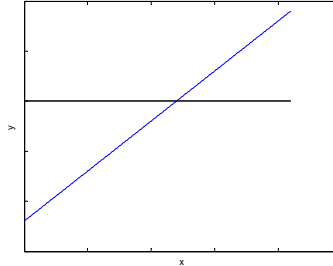


Esempio di soluzione non univoca



$$y = x + 2$$
$$2y = 2x + 4$$

$$A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$$



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

$$2 x_1 - 2 x_2 = -4$$

$$x_1 = x$$

$$x_2 = y$$

$$\det(A) = 1(-2) - (-1)(2) = -2 + 2 = 0$$

La soluzione non è unica: tutti i punti della retta soddisfano contemporaneamente le 2 equazioni



Risoluzione di un sistema 2x2



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

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

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

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



Sistema $M \times N$, $M > N$



$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1N}x_N &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2N}x_N &= b_2 \\ &\dots \\ a_{M1}x_1 + a_{M2}x_2 + \dots + a_{MN}x_N &= b_M \end{aligned}$$

Ammette 1, nessuna o ∞ soluzioni

$$A x = b$$

A è $M \times N$, $M > N$, non è una matrice quadrata.

1, nessuna, ∞ soluzioni.

Esempio:

$$\begin{aligned} 3x_1 + 2x_2 + \dots + 4x_N &= 5 \\ 4x_1 - 2x_2 + \dots + 0.5x_N &= 3 \\ &\dots \end{aligned}$$

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

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

Posso sempre calcolare la soluzione in forma matriciale.



Sistemi lineari con $m > n$

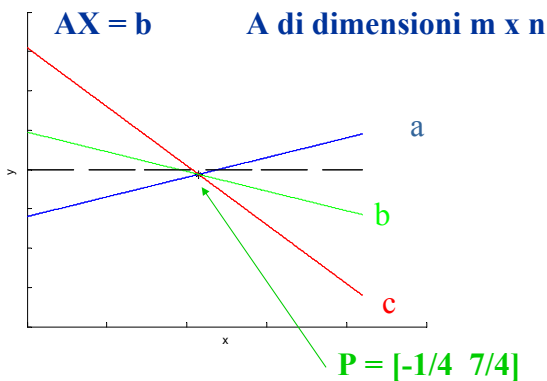


$J(W,L)$ è rettangolare: numero di righe maggiore del numero di colonne

$$\begin{aligned} y &= x + 2 \\ y &= -3x + 1 \\ y &= -x + 3/2 \end{aligned}$$

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

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



Esiste un'equazione "di troppo"

Nessuna, 1 o ∞ soluzioni

Rango di A è pieno



Rango di una matrice

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

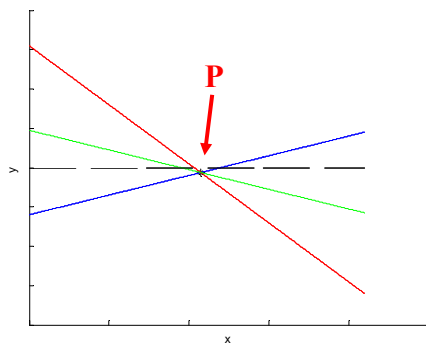
Data una matrice A di ordine n ($n \times n$),

una matrice A $n \times n$ 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 $n \times n$ ha rango n (rango pieno) se e solo se
il suo determinante è diverso da 0



Relazione tra le equazioni (combinazione lineare)



$$\begin{aligned} \alpha_1 (y - x - 2) + \\ \alpha_2 (y + 3x - 1) = \\ (y + x - 3/2) \end{aligned}$$

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:

$$A X = B \quad \longrightarrow \quad A' A X = A' B \quad \longrightarrow \quad (A' A)^{-1} A' A X = (A' A)^{-1} A' B$$

$$\downarrow$$

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

$$X = (A' A)^{-1} A' B$$

Quale criterio viene soddisfatto da X ?



Sistemi lineari con $m > n$



$$y = x - 2$$

$$y = -3x + 1$$

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

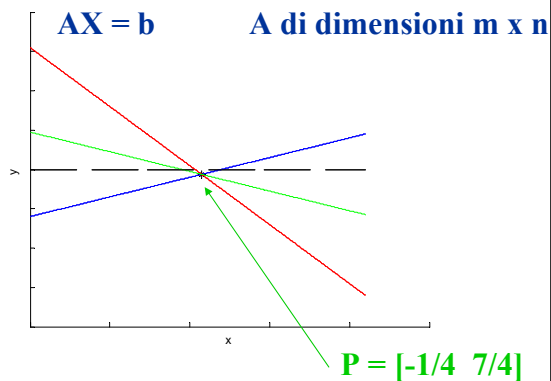
$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ -1 \\ +1.5 \end{bmatrix}$$


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

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


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

intersezione





Riformulazione del problema



$a_{11}x_1 + a_{12}x_2 + \dots + a_{1N}x_N = b_1 + v_1$
 $a_{21}x_1 + a_{22}x_2 + \dots + a_{2N}x_N = b_2 + v_2$

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

Modello **Misure**

$Ax = b + N$

Errore di modello (sistematico, randomico). $M \times 1 \Rightarrow$ **Residuo**.
 $M \times 1$
 Vettore dei termini noti


$M \times N$ $N \times 1$
 (Matrice di disegno) Vettore delle incognite

Quale criterio viene soddisfatto da X?


A.A. 2008-2009

27/59

<http://homes.dsi.unimi.it/~borghese/>



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 minima da tutte le rette.

$$\min_x \sum_k v_k^2 = \min_x (Ax - b)^2$$

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

$$A^T A x = A^T b$$

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

NB le funzioni costo sono spesso quadratiche (problemi di minimizzazione convessi) perchè il costo cresce sia che il modello sovrastimi che sottostimi le misure.

A.A. 2008-2009

28/59

<http://homes.dsi.unimi.it/~borghese/>



Sistemi lineari con $m > n$



$$\begin{aligned} y &= x - 2 \\ y &= -3x + 1 \\ y &= -x + 3/2 \end{aligned}$$

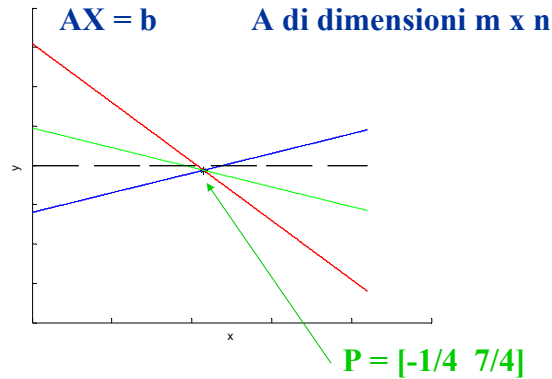
$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ -1 \\ -1.5 \end{bmatrix}$$

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

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

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

intersezione



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

A.A. 2008-2009

<http://homes.dsi.unimi.it/~borghese/>



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



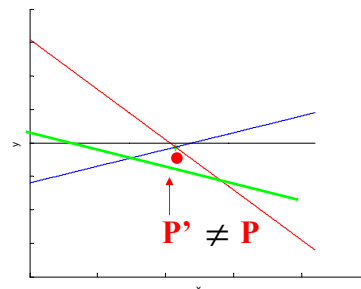
$$\begin{aligned} y &= x + 2 \\ y &= -3x + 1 \\ y &= -x + 1/2 \end{aligned}$$

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

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

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

$$AX = b \quad A \text{ di dimensioni } m \times n$$



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

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

No intersezione

A.A. 2008-2009

30/59

<http://homes.dsi.unimi.it/~borghese/>



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 dalla retta



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 parametri, dato un certo insieme di dati.

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**.



Sommario



Matrici

Sistemi lineari

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare



Condizionamento della matrice $C = A'A$



$X = (A'A)^{-1}A'B = CA'B$ - 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.

$$A x = b$$



Sistema lineare: soluzione robusta



$$A X = B \quad \longrightarrow \quad A' A X = A' B \quad \longrightarrow \quad X = (A' A)^{-1} A' B$$

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

Soluzione tramite Singular Value Decomposition (diagonalizzazione)

Numero di condizionamento varia circa con A .

$$A X = B \quad \longrightarrow \quad U W V X = B \quad \boxed{x = V' W^{-1} U' b}$$

Ortonormale $M \times N$ Diagonale ($N \times N$) Ortonormale $N \times N$

$$V^T W^{-1} U^T U W V X = V^T W^{-1} U^T B \quad \longrightarrow \quad X = V^T W^{-1} U^T B$$

- 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?

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

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 sovrapparametrizzato.



Valutazione della bontà della stima



$$x = (A^T A)^{-1} A^T b \iff \min_x \sum_k v_k^2 = \min_x (Ax - b)^2$$

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

$$\langle v_k \rangle = 0$$

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

Varianza della stima = varianza dell'errore di misura



Richiamo di statistica



Data una variabile casuale, x , il suo valore medio calcolato su N campioni è dato da:

$$M_x = \frac{\sum_{k=1}^N x_k}{N}$$

Data una variabile casuale, x , la sua varianza su N campioni è data da:

$$\sigma_x^2 = \frac{\sum_{k=1}^N (x_k - M_x)^2}{N}$$

Data una variabile casuale, x , la sua deviazione standard su N campioni è data da:

$$\sigma_x = \sqrt{\frac{\sum_{k=1}^N (x_k - M_x)^2}{N}}$$

Varianza e deviazione standard descrivono la dispersione attorno al valor medio.



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



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

$$x = C A' b$$

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

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.

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



$$u = C A' v$$

$$E[u] = 0$$



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:

$$\begin{bmatrix} u_1^2 & u_1 u_2 & \dots & u_1 u_W \\ u_2 u_1 & u_2^2 & \dots & u_2 u_W \\ \dots & \dots & \dots & \dots \\ u_W u_1 & u_W u_2 & \dots & u_W^2 \end{bmatrix}$$

$$\langle u_i, u_j \rangle$$

$$u = C A' v$$

\Rightarrow

$$u' = v' A (C)'$$

$u u' = C A' v v' A C' \Rightarrow$ Applicando l'operatore di media, si ottiene:

$$\langle u u' \rangle = C A' \langle v v' \rangle A C'$$

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 $\langle v v' \rangle = I \sigma_0^2$.



Correlazione tra i parametri

$$\langle uu' \rangle = CA' IA C' \sigma_0^2 = C' \sigma_0^2$$

$$\langle u'u \rangle = C \sigma_0^2$$

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.

$$-1 \leq r_{ij} = \frac{\langle u_i u_j \rangle}{\sqrt{\langle u_i \rangle^2 \langle u_j \rangle^2}} = \frac{c_{ij}}{\sqrt{c_i c_j}} \leq +1$$

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: $x = [x_1, x_2, \dots, x_N]$ 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:

$$C = \begin{bmatrix} \sigma_{x_1 x_1} & \sigma_{x_1 x_2} & \cdot & \sigma_{x_1 x_N} \\ \sigma_{x_2 x_1} & \sigma_{x_2 x_2} & \cdot & \sigma_{x_2 x_N} \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{x_N x_1} & \sigma_{x_N x_2} & \cdot & \sigma_{x_N x_N} \end{bmatrix}$$

Varianza: $\sigma_{x_i x_i} = \sigma_{x_i}^2$

N parametri

Covarianza: $\sigma_{x_i x_j} = \sigma_{x_j x_i} \quad i \neq j$

$(N-1)^2/2$ 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:

$$r = \frac{M_{x_i x_j} - M_{x_i} M_{x_j}}{\sigma_{x_i} \sigma_{x_j}} \quad -1 \leq r \leq +1$$

Definendo la covarianza tra x_i ed x_j come:

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

Dalla definizione di deviazione standard risulta:

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

A.A. 2008-2009

43/59

<http://homes.dsi.unimi.it/~borghese/>



Soluzione sistema lineare sovradeterminato



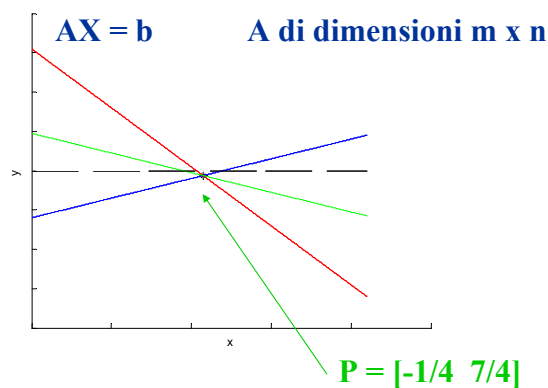
$$\begin{aligned} y &= x + 2 \\ y &= -3x + 1 \\ y &= -x + 3/2 \end{aligned}$$

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

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

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

$$P = C * A^T * b \quad P = [-0.25 \quad 1.75] \\ \text{intersezione}$$



A.A. 2008-2009

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

<http://homes.dsi.unimi.it/~borghese/>



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



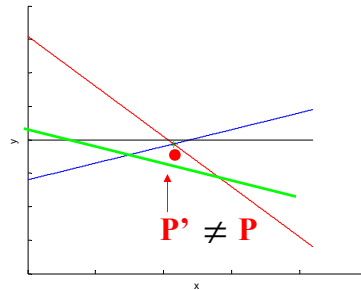
$$\begin{aligned} y &= x + 3 & +1 \\ y &= -3x + 1 & 0 \\ y &= -x + 1/2 & -1 \end{aligned}$$

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

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

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

$AX = b$ A di dimensioni $m \times n$



$$\sum_k v_k^2 = \|Ax - b\|^2 = 1.5$$

$$P = C * A^T * b \quad P' = [-0.5 \ +2.0]$$

No intersezione

A.A. 2008-2009

45/59

<http://homes.dsi.unimi.it/~borghese/>



Utilizzo di una terza retta diversa



$$\begin{aligned} y &= x + 2 \\ y &= -3x + 1 \\ y &= -8x - 1/4 \end{aligned}$$

$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -8 & -1 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ -1 \\ +1/4 \end{bmatrix}$$

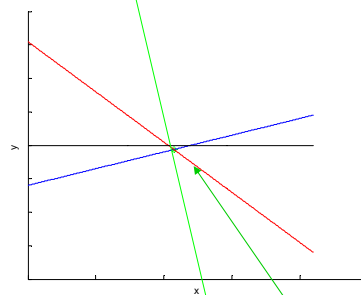
$$A^T * A = \begin{bmatrix} 74 & 10 \\ 10 & 3 \end{bmatrix}$$

$$\det(A^T * A) = 122$$

$$C = (A^T A)^{-1} = \begin{bmatrix} 0.0246 & -0.0820 \\ -0.0820 & 0.6066 \end{bmatrix} \quad P = C * A^T * b \quad P = [-0.25 \ +1.75]$$

intersezione

$AX = b$ A di dimensioni $m \times n$



A.A. 2008-2009

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

<http://homes.dsi.unimi.it/~borghese/>



Utilizzo di una terza retta diversa Stima variabilità di P

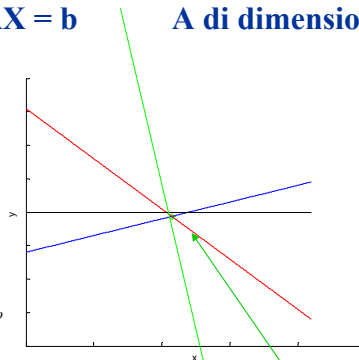


$$\begin{aligned} y &= x + 2 \\ y &= -3x + 1 \\ y &= -8x - 1/4 \end{aligned}$$

$$AX = b \quad A \text{ di dimensioni } m \times n$$

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

$$\langle uu' \rangle = \begin{bmatrix} 0.0246 & -0.0820 \\ -0.0820 & 0.6066 \end{bmatrix} \sigma^2_o$$



$$P = [-1/4 \quad 7/4]$$

Stimo ancora meglio x_1 di x_2 ,
cioè la variabile x rispetto alla y .

$$P = [-0.25 \quad +1.75]$$

intersezione

A.A. 2008-2009

47/59

<http://homes.dsi.unimi.it/~borghese/>



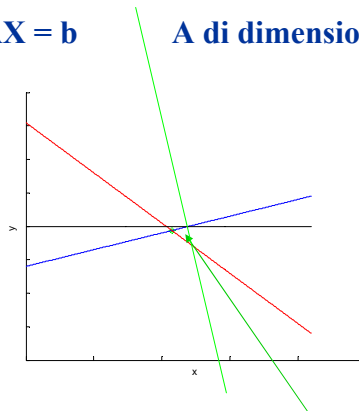
Infatti se ho un errore di misura...



$$\begin{aligned} y &= x + 3 & +1 \\ y &= -3x + 2 & 0 \\ y &= -8x - 5/4 & -1 \end{aligned}$$

$$AX = b \quad A \text{ di dimensioni } m \times n$$

$$b = \begin{bmatrix} -3 \\ -1 \\ +5/4 \end{bmatrix}$$



$$P = C * A^T * b \quad P = [-0.4795 \quad 2.8484]$$

intersezione

Stimo meglio x_2 di x_1 , cioè errore
nella posizione della retta influisce
principalmente su y e poco su x .

$$\|Ax - b\| \neq 0 = 0.5246$$

A.A. 2008-2009

48/59

<http://homes.dsi.unimi.it/~borghese/>



Esempio

Supponiamo di avere distorsione radiale misurata come segue:

Distanza [mm]	Distorsione radiale [mm]
0.000	0.000
20.072	0.004
40.855	0.018
63.155	0.047
88.034	0.062
116.995	0.035
152.472	-0.064

Modello lineare utilizzato: $\Delta r = k_1 r + k_2 r^3 + k_3 r^5 + k_4 r^7$.

Parametri determinati:

$k_1 = 1.99697 \times 10^{-4}$
$k_2 = 1.94801 \times 10^{-7}$
$k_3 = -1.97388 \times 10^{-11}$
$k_4 = 0.43934 \times 10^{-15}$

Analisi di varianza?



Sommario

Sistemi lineari e matrici

Soluzione dei sistemi lineari

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, $f(\mathbf{P}; \mathbf{W}) = 0$, la linearizzazione si può scrivere come:

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

E' un'equazione lineare nei dw che descrive il comportamento della funzione $F(\cdot)$ nell'intorno del punto P_o con i parametri W_o .



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.
- 3) Correzione dei parametri.

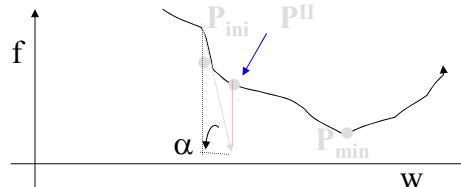
Può essere pesante perchè richiede l'inversione della matrice di covarianza.
Spesso si preferiscono utilizzare metodi di ottimizzazione del primo ordine.



Minimizzazione tramite gradiente: 1 variabile



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



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.

$$dw \propto -f'(w;P) \quad \text{dati } P, w.$$

Occorre un'inizializzazione.

Metodo iterativo.

esel



Esempio di applicazione tecnica del gradiente per funzioni di 1 variabile



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

Misuriamo un punto sulla parabola: $x = 1, y = 3$.

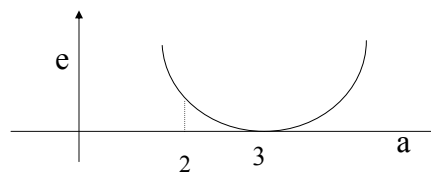
Vogliamo modificare a in modo che la parabola passi per $P(x,y)$.

La funzione costo da minimizzare sarà: $e = f(a | x,y) = (y - ax^2)^2$

La soluzione è $a = 3$

Partiamo da $a_{ini} = 2$.

$$e = (3 - 2 \cdot 1)^2 = 1$$



Utilizziamo il metodo del gradiente:

Calcoliamo la derivata di $f(a | x,y) \rightarrow f'(a) = -2(y - ax^2)x^2$



Minimizzazione - underdamping



Consideriamo $\alpha = 1$

Calcoliamo la derivata di $f(.) \rightarrow f'(.) = -2 (y - a x^2) x^2$

Utilizziamo il metodo del gradiente:

Passo 1:

Calcoliamo l'incremento da dare al parametro a:

$$da = -[-2 (3 - 2 \cdot 1) \cdot 1] = -[-6 + 4] = 2 \quad a' = 2 + 2 = 4$$

Passo 2:

Calcoliamo l'incremento da dare al parametro a:

$$da = -[-2 (3 - 4 \cdot 1) \cdot 1] = -[-6 + 8] = -2 \quad a'' = 4 - 2 = 2$$

Oscillazioni!!!

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



Minimizzazione -2 passi



Consideriamo $\alpha = 0.4$

Calcoliamo la derivata di $f(.) \rightarrow f'(.) = -2 (y - a x^2) x^2$

Utilizziamo il metodo del gradiente:

Passo 1:

Calcoliamo l'incremento da dare al parametro a:

$$da = -0.4 [-2 (3 - 2 \cdot 1) \cdot 1] = -[-6 + 4] = 0.8 \quad a' = 2 + 0.8 = 2.8$$

Passo 2:

Calcoliamo l'incremento da dare al parametro a:

$$da = -0.4 [-2 (3 - 2.8 \cdot 1) \cdot 1] = -[-6 + 5.6] = 0.16 \quad a'' = 2.8 + 0.16 = 2.96$$

Converge ad $a = 3$.

Posso correre il rischio di spostarmi troppo lentamente



Minimizzazione di funzioni di più variabili



$\min(f(\mathbf{x}, \mathbf{w}))$ funzione costo od errore, \mathbf{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, $f(\cdot)$.
Estensione della tecnica del gradiente a più variabili.

$$d\mathbf{w} = -\alpha \nabla f(\mathbf{x}; \mathbf{w}), \text{ dato } \mathbf{P}, \mathbf{W}.$$

Serve un'approssimazione iniziale per i pesi $\mathbf{W}_{\text{ini}} = \{w_j\}_{\text{ini}}$.



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.



Sommario



Sistemi lineari e matrici

Soluzione dei sistemi lineari

Analisi dell'affidabilità della stima

Linearizzazione di sistemi di funzioni

Determinazione dei parametri di un modello non-lineare.