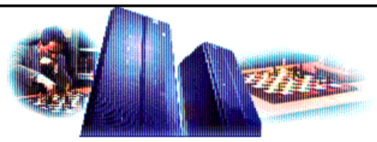
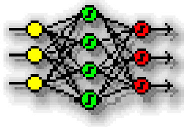




Sistemi Intelligenti
I sistemi lineari

Alberto Borghese

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

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

Sistema lineare

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

....

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

{aij} – coefficienti in numero N x M
{xj} – incognite, M
{bj} – termini noti, N

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

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

....

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

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 ≠ BA

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

$$AI = A = IA$$

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

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

Determinante di una matrice Quadrata

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

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

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

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

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

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

Matrice Inversa

A = [$a_{i,j}$], **matrice quadrata.**

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

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

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

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

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

Se esiste, la matrice inversa è unica.

Altre proprietà delle matrici

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

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

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

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

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

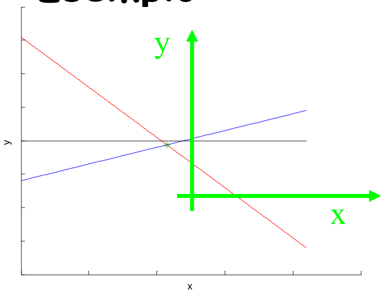
Condizione di ortonormalità:
Il determinante è = 1.
La somma dei prodotti di due righe o di due colonne è = 0.
La somma dei quadrati degli elementi su righe e colonne = 1
Esempio notevole: **matrice di rotazione (cambio di sistema di riferimento)**.

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

Esempio

$y = x + 2$

$y = -3x + 1$



$x_1 = x$

$x_2 = y$

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

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

Risolvo per sostituzione:

$-3\ (-2 + x2) - x2 = -1 \;\Rightarrow\; x2 = 7/4$

$x_1 = -2 + 7/4 \;\Rightarrow\; x1 = -1/4$

Soluzione dei sistemi lineari

$y = x + 2$

$y = -3x + 1$

Scrivo il sistema lineare: $Ax = b$

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

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

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

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).

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

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

....

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

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

M x N
(Matrice di disegno)

N x 1
Vettore delle incognite

M x 1
Vettore dei termini noti

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

....

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

Ammette 1, nessuna o ∞ soluzioni

A x = **b**

A è N x N quadrata

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

altrimenti, nessuna (rette parallele)
o
∞ soluzioni (rette coincidenti).

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.

1, nessuna, ∞ soluzioni.

Ho delle equazioni di troppo, devono essere correlate, combinate linearmente).

Posso sempre calcolare la soluzione in forma matriciale.

Sistema lineare: soluzione algebrica

Caso generale:

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

$$\Downarrow$$

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

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

Quale criterio viene soddisfatto da X?

Riformulazione del problema

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

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

....

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

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

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.

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

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

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

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

Giustificazione statistica

http://www.weibull.com/AccelTestWeb/mle_maximum_likelihood_parameter_estimation.htm

Data set, S, di N punti (**a**, b) da fittare con un modello in M parametri, {x}. b = f(**a**;**x**).

Quale relazione c'è tra la minimizzazione dell'errore residuo e la statistica?

Dato un certo insieme di parametri, vogliamo valutare quanto sia verosimilile estrarre il data set S. Ovverosia la *likelihood* (verosimiglianza) dei parametri, dati i dati S ➔ Parametri tali che la likelihood sia massimizzata.

Ogni punto viene visto come estratto da una distribuzione statistica parametrica:

$$f(\mathbf{a}, b; x_1, x_2, \ldots x_N)$$

Dove i {θ} sono non noti e devono essere determinati.

Funzione di verosimiglianza: $L(\mathbf{x} \mid \theta) =$

$$L(a_1, a_2, \ldots a_M, b; x_1, x_2, \ldots x_N) = \prod_{i=1}^{N} f(a_i, b_i; x_1, x_2, \ldots x_N)$$

Massimizzazione

$$L(a_1, a_2, \ldots a_M, b; x_1, x_2, \ldots x_N) = \prod_{i=1}^{N} f(a_i, b_i; x_1, x_2, \ldots x_N)$$

Si passa al logaritmo, il prodotto diventa una somma.

$$\Lambda = \ln L = \sum_{i=1}^{N} \ln\left(f\left(\mathbf{a}, b; x_1, x_2 \ldots x_N\right)\right)$$

Massimizzazione rispetto ai {θ}

$$L'(\mathbf{x}) / L(\mathbf{x}) = \frac{\partial \Lambda}{\partial x_j} = 0, \; j = 1, 2, 3, \ldots N$$

Sistema di tante equazioni quanti sono i parametri.

Dalla massima verosimiglianza ai minimi quadrati - I

Supponiamo l'errore sui dati Gaussiano: $f(T) = \frac{1}{\sigma_T \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{T-\bar{T}}{\sigma_T}\right)^2}$

$$L(T_1, T_2, \ldots, T_N | \bar{T}, \sigma_T) = L = \prod_{i=1}^{N}\left[\frac{1}{\sigma_T\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{T_i-\bar{T}}{\sigma_T}\right)^2}\right]$$

$$L = \frac{1}{(\sigma_T\sqrt{2\pi})^N} e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{T_i-\bar{T}}{\sigma_T}\right)^2}$$

Scriviamo Λ:

$$\Lambda = \ln(L) = -\frac{N}{2}\ln(2\pi) - N\ln\sigma_T - \ln\left(\frac{1}{2}\sum_{i=1}^{N}\left(\frac{T_i - \bar{T}}{\sigma_T}\right)^2\right)$$

Dalla massima verosimiglianza ai minimi quadrati - II

Then taking the partial derivatives of with respect to each one of the parameters and setting it equal to zero yields:

$$\frac{\partial(\Lambda)}{\partial \bar{T}} = \frac{1}{\sigma_T^2}\sum_{i=1}^{N}(T_i - \bar{T}) = 0 \qquad \frac{\partial(\Lambda)}{\partial \sigma_T} = -\frac{N}{\sigma_T} + \frac{1}{\sigma_T^3}\sum_{i=1}^{N}(T_i - \bar{T})^2 = 0$$

Risolvendole simultaneamente, si ottiene:

$$\bar{T} = \frac{1}{N}\sum_{i=1}^{N} T_i \qquad \hat{\sigma}_T^2 = \frac{1}{N}\sum_{i=1}^{N}(T_i - \bar{T})^2 \qquad \hat{\sigma}_T = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - \bar{T})^2}$$

Stima ai minimi quadrati e MaxVer

$$\frac{\partial(\Lambda)}{\partial \bar{T}} = \frac{1}{\sigma_T^2}\sum_{i=1}^{N}(T_i - \bar{T}) = 0 \qquad \frac{\partial(\Lambda)}{\partial \sigma_T} = -\frac{N}{\sigma_T} + \frac{1}{\sigma_T^3}\sum_{i=1}^{N}(T_i - \bar{T})^2 = 0$$

$$\bar{T} = \frac{1}{N}\sum_{i=1}^{N} T_i$$

Se supponiamo che il modello, dal quale viene calcolato il residuo, non sia rappresentato da un singolo parametro, ma da una combinazione lineare dei parametri, otteniamo:

$$\vartheta = \sum_{k=1}^{N} a_{ij} x_j \qquad \min \frac{1}{\sigma^2}\sum_{j=1}^{M}\left(T_j - \sum_{i=1}^{N} a_{ij}\vartheta_i\right)^2$$

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

Condizionamento della matrice C = A'*A

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

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

Sistema lineare: soluzione robusta

A X = B ⟹ **A' A X = A' B** ⟹ **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

U W V X = B

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

Ortonormale M x N

Diagonale (N x N)

Ortonormale N x N

$V^T W^{-1} U^T U W V X = V^T W^{-1} U^T B$ ➔ $X = V^T W^{-1} U^T B$

- 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?

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

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

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

Si può facilmente osservare valutando il valore singolare più piccolo della matrice W.

In questo caso il problema è sovraparametrizzato.

Valutazione della bontà della stima

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

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

$<v_k> = 0$

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

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:

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

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 = CA'* b

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

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.

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

⇓

u = C A'* v

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:

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

$<u_i, u_j>$

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

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

<uu'> = C A' <vv'>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 $<vv'>= I\sigma_0^2$.

Correlazione tra i parametri

$<uu'> = CA'\ IA\ C'\sigma_0^2 = C'\ \sigma_0^2$ $\boxed{<u'u> = C\ \sigma_0^2}$

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,… 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:

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

$-1 \leq r \leq +1$

Definendo la covarianza tra x_i ed x_j come:

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

Dalla definizione di deviazione standard risulta:

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

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

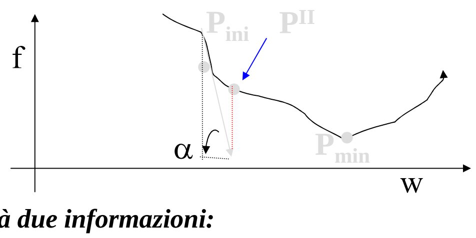
Analisi dell'affidabilità della stima

Linearizzazione di sistemi di funzioni

Determinazione dei parametri di un modello non-lineare

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.

$dw \propto -f'(w;P)$ dati P, w.

Occorre un'inizializzazione.

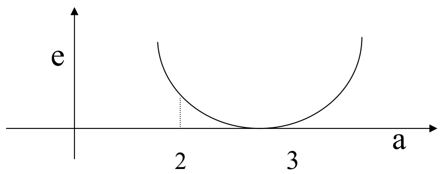
Metodo iterativo.

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: $e = f(x,y \mid w) = (y-ax^2)^2$
La soluzione è a = 3

Partiamo da $a_{ini} = 2$.



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

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 (e.g. robotica)

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 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 $W_{ini} = \{w_j\}_{ini}$.

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:

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

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

Stima di parametri in insiemi di equazioni non lineari - sistema

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

Jacobiano:

$$J = \begin{bmatrix} \left.\frac{\partial F_1(.)}{dx_1}\right|_{W_o,P_o} & \left.\frac{\partial F_1(.)}{dx_2}\right|_{W_o,P_o} & \left.\frac{\partial F_1(.)}{dx_N}\right|_{W_o,P_o} \\ \left.\frac{\partial F_2(.)}{dx_1}\right|_{W_o,P_o} & \left.\frac{\partial F_2(.)}{dx_2}\right|_{W_o,P_o} & \left.\frac{\partial F_2(.)}{dx_N}\right|_{W_o,P_o} \\ .... & & \\ \left.\frac{\partial F_M(.)}{dx_1}\right|_{W_o,P_o} & & \left.\frac{\partial F_M(.)}{dx_N}\right|_{W_o,P_o} \end{bmatrix}$$

F(**P;W**) = F(**P$_o$;W$_o$**) + **J dw**

Sistemi di equazioni lineari in **dw** (vettore)

Soluzione del problema di ottimizzazione non lineare

$\min(f(\mathbf{x}; \mathbf{w})) = \min (F^*)^2$

Il minimo della funzione F* sarà ≠ 0 (errore di modellazione + errore di misura).

Inizializzazione
Determino un valore dei parametri **w** da cui partire.

Linearizzazione (sviluppo di Taylor del primo ordine) attorno al valore attuale dei parametri

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

Soluzione del sistema lineare nei dw_j

Update
Aggiorno il valore dei parametri w

Sommario

Sistemi lineari e matrici

Soluzione dei sistemi lineari

Analisi dell'affidabilità della stima

Linearizzazione di sistemi di funzioni

Determinazione dei parmetri di un modello non-lineare.