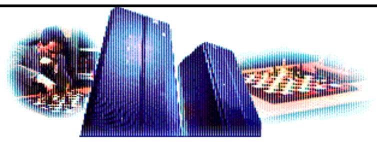
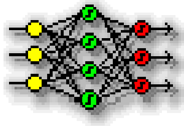




Sistemi Intelligenti
I sistemi lineari

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

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

....

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

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

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

....

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

Esempio

Barca a vela con motore + vento

3 x1 + 5 x2 = 100

Il vento produce un effetto di 3 volte la sua velocità.
Il motore produce un effetto di 5 volte la sua velocità di rotazione.

Se x2 = 10, x1 = 25, sono collegati. Quanto valgono in quella particolare configurazione?

2 x1 + 2 x2 = 50

Manteniamo il numero di giri del motore ed il vento, ma il mare ingrossa ed il rendimento dei due motori cambia.

Ottengo x2 ≅ 16.6   x1 ≅ 8.3

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)

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

Matrici (Proprietà)

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

Il prodotto è associativo rispetto alla somma ma non commutativo (A+B)C = AC + BC.
AB ≠ BA

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

$$AI = A = IA$$

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

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

Determinante

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

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

A* minore complementare di A

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

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 = $[a_{i,j}]$

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

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

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

Matrici ortonormali

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

Condizione di ortogonalità:

- La somma dei prodotti di due righe o di due colonne è = 0.

Condizione di normalità:

- Il determinante è = 1.

Condizione di normalità:

- Il determinante è = 1.
- La somma dei prodotti di due righe o di due colonne è = 0.

$\Rightarrow$ La somma dei quadrati degli elementi di una riga o colonna è = 1.

$\Rightarrow$ $A' = A^{-1}$

Sistema lineare

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

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

....

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

A x = b

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

M x 1

Vettore dei termini noti

Esempio:

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

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

....

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

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 N x N

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

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

....

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

Esempio:

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

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

....

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

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

A è N x N quadrata

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

1, ∞ soluzioni.

Sistema M x N, M > N

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

....

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

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

A è M x N, M > N.

1, nessuna, ∞ soluzioni.

Ho delle equazioni di troppo.
Scelgo N delle M righe e scrivo A*

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

Sistema lineare: soluzione algebrica

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

$\Downarrow$

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

Quale criterio viene soddisfatto da X?

Rappresentazione statistica

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

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

$\ldots\ldots\ldots\ldots$

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

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 minimo

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

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

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

Esempio:

$(a_{11}\,x_1 + a_{12}\,x_2 - b_1)^2$

$R = (a^2_{11}\,x^2_1 + a^2_{12}\,x^2_2 + b^2_1\,a^2_{11} \quad + \quad 2\,a_{11}\,a_{12}\,x_1\,x_2 - 2\,a_{11}\,b_1\,x_1 + a_{12}\,b_1\,x_2)$

$$\frac{\partial}{\partial x_1} R = \quad 2\,a^2_{11}\,x_1 + \quad 2\,a_{11}\,a_{12}\,x_2 - 2\,a_{11}\,b_1\,x_1\,)$$

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:

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

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

Funzione di verosimiglianza: L(**x** | θ) =

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

Massimizzazione

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

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}

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

Sistema di tante equazioni quanti sono i parametri.

Dalla massima verosimiglianza ai minimi quadrati - I

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

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

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

Scriviamo Λ:

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

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:

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

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

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

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

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.

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

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

Numero di condizionamento varia circa con A.

A X = B

U W V X = B

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

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

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

- 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

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 e(0,σ^2)

$<v_k> = 0$

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

.

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.

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

$$\Downarrow$$

u = C A'* v

Impostazione del calcolo della correlazione tra i parametri

u = C A' v

Vogliamo individuare la correlazione tra due parametri r ed s. Devo quindi determinare il valore atteso di $u_r * u_s$.

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

u = C $^{-1}$A' v => u' = v'A (C $^{-1}$)'

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'>= $I\sigma_0^2$.

Calcolo della correlazione tra i parametri

$<uu'> = CA' \, IA \, C' \sigma_0^2 = C \, \sigma_0^2$

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

Segue che: $\sigma^2(u_{ij}) = c_{ij}^{-1}\sigma_0^2$ 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.

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

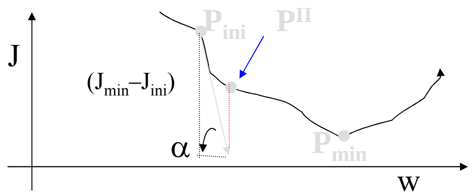
Linearizzazione di sistemi di funzioni

Determinazione dei parametri di un modello non-lineare

Minimizzazione tramite gradiente

Minimizzo J(.) rispetto ai parametri.

Tecnica del gradiente applicata alla minimizzazione di funzioni non-lineari di **una variabile**: J= J(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.Mi muovo lungo la direzione della pendenza, arrivo in P^{II}. Calcolo $J(w^{II})$.

Da qui riparto fino a quando non arrivo in P_{min}.

Minimizzazione di funzioni di più variabili

min(J{**w**}| …..) funzione costo od errore

Gradiente:
$$\frac{\partial J(\{w\}|....)}{\partial w_1}\frac{\boldsymbol{w_1}}{|\boldsymbol{w_1}|}+\frac{\partial J(\{w\}|....)}{\partial w_2}\frac{\boldsymbol{w_2}}{|\boldsymbol{w_2}|}+$$
$$\frac{\partial J(\{w\}|....)}{\partial w_3}\frac{\boldsymbol{w_3}}{|\boldsymbol{w_3}|}+\frac{\partial J(\{w\}|....)}{\partial w_4}\frac{\boldsymbol{w_4}}{|\boldsymbol{w_4}|}+........$$

Modifico il valore dei pesi di una quantità proporzionale alla pendenza della funzione costo rispetto a quel parametro. Estensione della tecnica del gradiente a più variabili.

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

Esempio di applicazione tecnica del gradiente per funzioni di 1 variabile

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

Consideriamo di avere stimato a = 2 e di avere misurato y = 3 per x = 1. Possiamo modificare il parametro a in modo tale che il modello fitta i dati (x=1 e y =3).

Linearizziamo il modello e scriviamo: $(y - ax^2) = \left[\frac{d}{da}(ax^2)\right]da$

Sostituendo i valori otteniamo: da = (3 – 2 · 1) / (1) = 1

a = 2+1 = 3. da = (3 – 3 · 1) / 1 = 0

Avendo più di una misurazione dobbiamo "mediare" tra gli incrementi dx calcolati per ogni misurazione.

Minimizzazione di funzioni di più variabili

min(J{**w**}| …..) funzione costo od errore

Gradiente:
$$\frac{\partial J(\{w\}|....)}{\partial w_1}\frac{\boldsymbol{w_1}}{|\boldsymbol{w_1}|}+\frac{\partial J(\{w\}|....)}{\partial w_2}\frac{\boldsymbol{w_2}}{|\boldsymbol{w_2}|}+$$
$$\frac{\partial J(\{w\}|....)}{\partial w_3}\frac{\boldsymbol{w_3}}{|\boldsymbol{w_3}|}+\frac{\partial J(\{w\}|....)}{\partial w_4}\frac{\boldsymbol{w_4}}{|\boldsymbol{w_4}|}+........$$

Modifico il valore dei pesi di una quantità proporzionale alla pendenza della funzione costo rispetto a quel parametro. Estensione della tecnica del gradiente a più variabili.

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

Minimizzazione dei residui (problema di minimo non lineare)

min(F(P|{x})-b)^2 = min (F*)^2

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

Inizializzazione
Determino un valore dei parametri x da cui partire.

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

$$F^*(P) = F^*(P_o) + \sum_{j=1}^{W} \left.\frac{\partial F^*(.)}{\partial x_j}\right|_{P_o} * dx_j = \mathbf{b - Adx}$$

Soluzione del sistema lineare nei dx_j

Update
Aggiorno il valore dei parametri x

Esempio

$F(P;x) = w_1\ G(P;c_1,\sigma_1) + w_2\ G(P;c_2,\sigma_2)$

Dato un certo insieme di campioni, quanto devono valere i parametri per ottenere un buon fitting?

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.