

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 Scienze dell'Informazione
borghese@dsi.unimi.it



A.A. 2010-2011

1/62

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



Sommario



Matrici e Sistemi lineari

I sistemi

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare

A.A. 2010-2011

2/62

<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 : $\bar{v} = \bar{u}^T M$



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 ($n \times n$),

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.



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)**.



Sommario



Matrici e Sistemi lineari

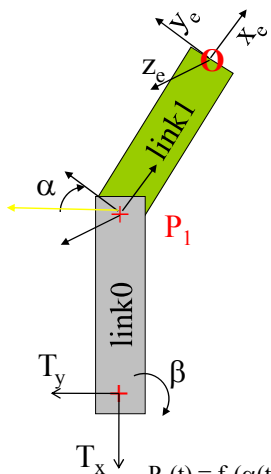
I sistemi

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare



Esempio di "sistema"



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

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

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



Le equazioni del sistema

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

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

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

$b = J X$

A.A. 2010-2011 11/62 <http://homes.dsi.unimi.it/~borghese/>



Sommario

Matrici

Sistemi lineari

Analisi dell'affidabilità della soluzione

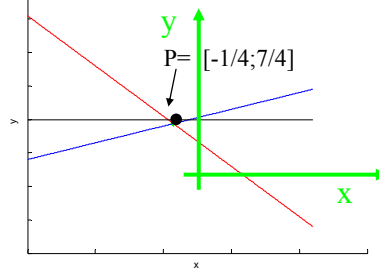
Determinazione dei parametri di un modello non-lineare



Esempio

$$y = x + 2$$

$$y = -3x + 1$$



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

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

$$y = x_2$$

$$x = x_1$$

Risolve 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. 2010-2011

13/62

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



Sistema lineare

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

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

.....

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

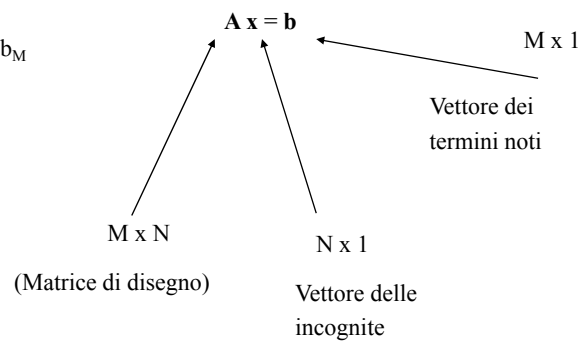
Esempio:

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

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

.....

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



A.A. 2010-2011

14/62

<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

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

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

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

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

Esempio:

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

altrimenti, **nessuna** (rette parallele)

o

∞ **soluzioni** (rette coincidenti).

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



Soluzione dei sistemi lineari



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

$$y = x + 2$$

$$y = -3x + 1$$

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

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

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

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

A.A. 2010-2011

17/62

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

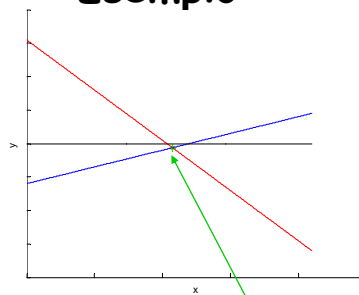


Esempio



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

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



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

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

$$x_1 = x$$

$$x_2 = y$$

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

Rango di A è pieno

$$x_1 = -1/4$$

$$x_2 = 7/4$$

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

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

A.A. 2010-2011

18/62

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

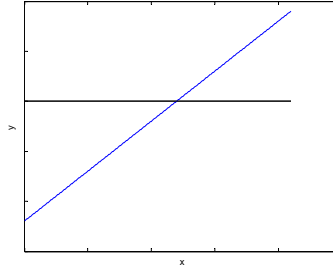


Esempio di soluzione non univoca



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

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



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

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

$$x_1 = x$$

$$x_2 = y$$

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

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



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

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

.....

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

Ammette 1, nessuna o ∞ soluzioni

$$A x = b$$

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

1, nessuna, ∞ soluzioni.

Esempio:

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

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

.....

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

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

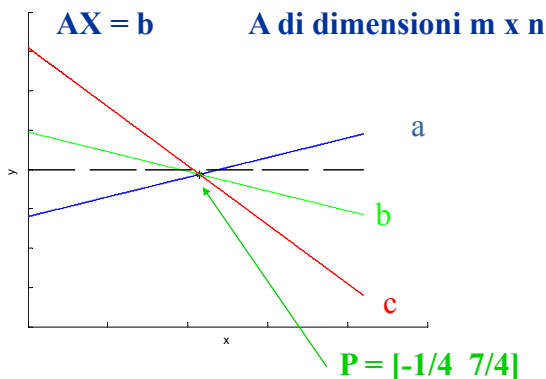
$$y = x + 2$$

$$y = -3x + 1$$

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

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

$\det(A^{sij})$ 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

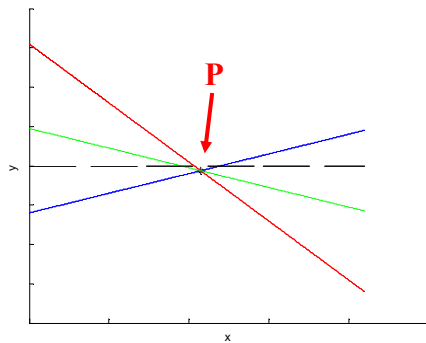
A.A. 2010-2011

23/62

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



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.

A.A. 2010-2011

24/62

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



Sistema lineare: soluzione algebrica

Caso generale:

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

\Downarrow

$(\mathbf{A}'\mathbf{A})$ gioca il ruolo di \mathbf{A} quadrata.

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

Quale criterio viene soddisfatto da \mathbf{X} ?



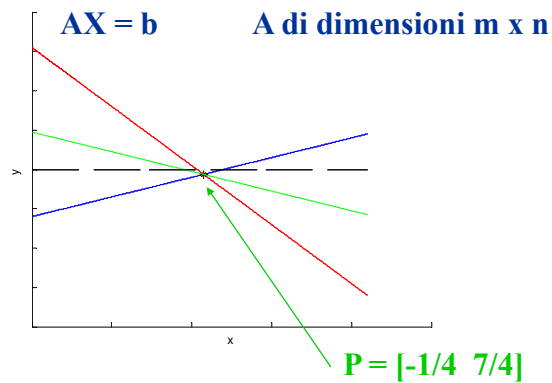
Sistemi lineari con $m > n$

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

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

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

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



$$\mathbf{P} = \mathbf{C} * \mathbf{A}^T * \mathbf{b} \quad \mathbf{P} = \begin{bmatrix} -0.25 & 1.75 \end{bmatrix}$$

intersezione

Riformulazione del problema con rumore

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

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

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

$Ax = b + N$

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

$M \times 1$ Vettore dei termini noti

Modello **Misure**

Quale criterio viene soddisfatto da X?

A.A. 2010-2011 27/62 <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.

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

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

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

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

A.A. 2010-2011 28/62 <http://homes.dsi.unimi.it/~borghese/>



Sistemi lineari con $m > n$



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

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

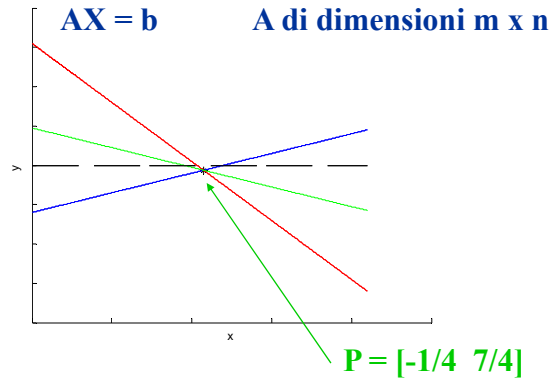
$$A^T * A = \begin{bmatrix} 11 & 3 \\ 3 & 3 \end{bmatrix} \quad \det = 24$$

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

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

intersezione

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



A.A. 2010-2011

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



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



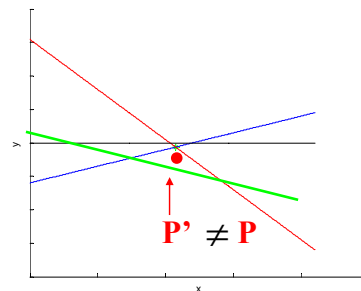
$$\begin{aligned} y &= x + 2 \\ y &= -3x + 1 \\ y &= -x + 1/2 \end{aligned}$$

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

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

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

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



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

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

No intersezione

A.A. 2010-2011

30/62

<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



Soluzione del sistema semplificato



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

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

$$\begin{bmatrix} \Delta x_e \\ \Delta y_e \\ 0 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 0 & -2,5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta \alpha \\ \Delta \beta \\ \Delta T_x \\ \Delta T_y \end{bmatrix}$$

$b = A X$



Condizionamento della matrice $C = A^*A$



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

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 .

$$A X = B$$

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

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

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

- 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?

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

$$x = V^*W^{-1}U^*b$$

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.



Sommario



Matrici

Sistemi lineari

Analisi dell'affidabilità della stima

Determinazione dei parametri di un modello non-lineare



Le distribuzioni statistiche



- Data una certa misura, questa può assumere valori diversi con frequenze diverse. La curva che descrive questi valori diversi si chiama curva di densità di probabilità.

A probability density function is most commonly associated with continuous univariate distributions. A random variable X has density f , where f is a non-negative Lebesgue-integrable function, if (Wikipedia):

$$P[a \leq X \leq b] = \int_a^b f(x) dx$$



I momenti di una variabile 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:

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

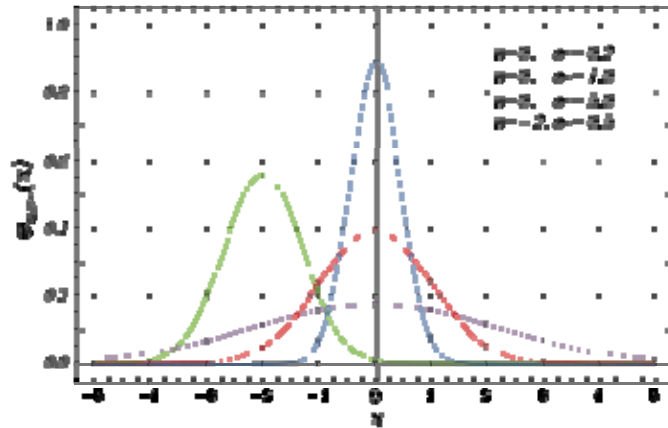
Data una variabile casuale, x , la sua deviazione standard su N campioni è data da:

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

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



Distribuzioni notevoli: Gaussiana



The probability that the random variable X lies in an interval whose width is related with the standard deviation, is

$$\Pr\{|X - \mu| \leq 1\} = 2 \cdot \text{erf}(1) = 0.68268 \quad (1.6)$$

$$\Pr\{|X - \mu| \leq 2\} = 2 \cdot \text{erf}(2) = 0.95452 \quad (1.7)$$

$$\Pr\{|X - \mu| \leq 3\} = 2 \cdot \text{erf}(3) = 0.9973 \quad (1.8)$$

$$p(x) = \frac{1}{\sqrt{(2\pi\sigma^2)^p}} e^{-\frac{|x-\mu|^2}{\sigma^2}}$$

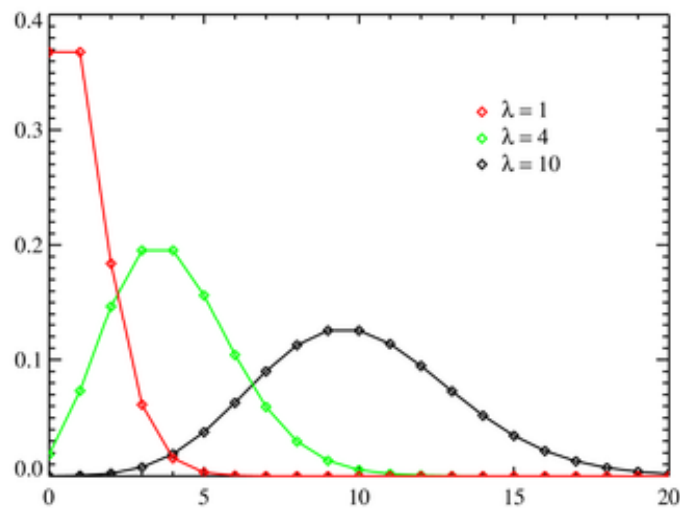
A.A. 2010-2011

39/62

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



Poisson distribution



A.A. 2010-2011

40/62

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



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**.



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



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:

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



Caso 2D



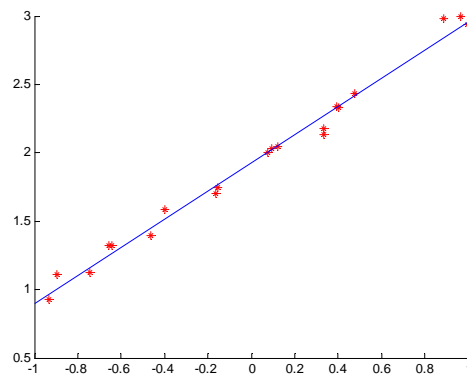
$N = 20$ punti $\sigma_0^2 = 0.01$
m reale = 1 q reale = 2

$$C = \begin{pmatrix} 0.1427 & -0.0002 \\ -0.0002 & 0.0500 \end{pmatrix}$$

m stimato = 1.0302
q stimato = 1.9308

$$C = \begin{pmatrix} 0.1702 & 0.0124 \\ 0.0124 & 0.0509 \end{pmatrix}$$

m stimato = 0.9937
q stimato = 1.9522



$$y = mx + q$$



Caso 2D - less points



$N = 10$ punti $\sigma_0^2 = 0.01$
 m reale = 1 q reale = 2

C =
 0.5927 -0.0030
 -0.0030 0.1000

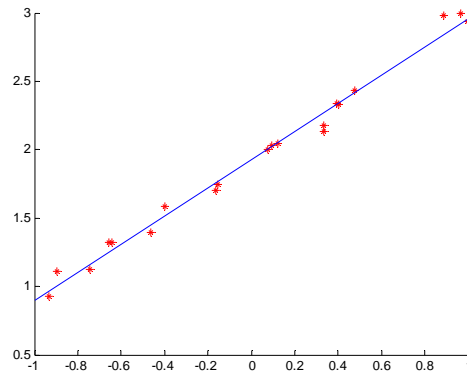
m_stimato =
 1.0081

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 $\sigma_0^2 = 0.01$
 m reale = 1 q reale = 2

C =
 0.0327 -0.0034
 -0.0034 0.0103

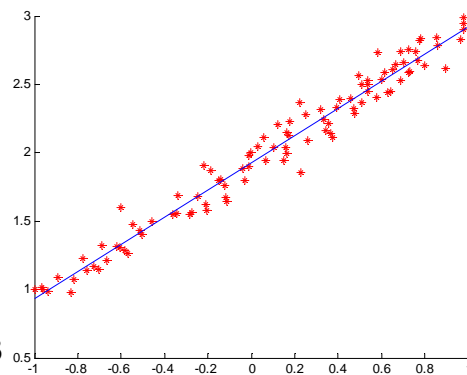
m_stimato =
 0.9935

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



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:

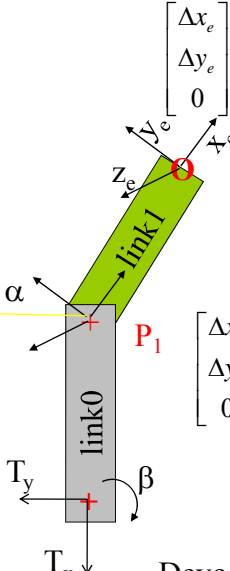
$$F(\mathbf{P}; \mathbf{W}) = F(\mathbf{P}_o; \mathbf{W}_o) + \sum_{j=1}^w \left. \frac{\partial F(\cdot)}{\partial w_j} \right|_{\mathbf{P}_o, \mathbf{W}_o} * dw_j = k - \sum_{j=1}^w a_j * dw_j$$

E' un'equazione lineare nei dw che descrive il comportamento della funzione $F(\cdot)$ nell'intorno del punto \mathbf{P}_o con i parametri \mathbf{W}_o .



Soluzione del sistema semplificato






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

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

$$\begin{bmatrix} \Delta x_e \\ \Delta y_e \\ 0 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 0 & -2,5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta \alpha \\ \Delta \beta \\ \Delta T_x \\ \Delta T_y \end{bmatrix}$$


$b = J X$




A.A. 2010-2011

Devo determinare 53/62

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



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.

A.A. 2010-2011

54/62

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



Le equazioni del sistema

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

$$\begin{bmatrix} \Delta \alpha \\ \Delta \beta \\ \Delta T_x \\ \Delta T_y \end{bmatrix} = (J^T J)^{-1} J^T \begin{bmatrix} \Delta x_e \\ \Delta y_e \\ 0 \end{bmatrix}$$

A.A. 2010-2011

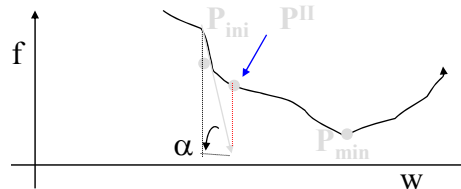
55/62

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



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

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



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.

ese\



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

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

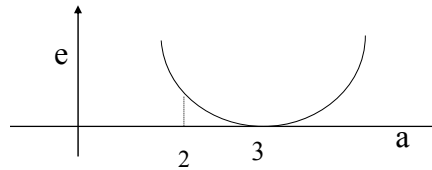
Vogliamo modificare a in modo che la parabola passi per $P(x,y)$.

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

La soluzione è $a = 3$

Partiamo da $a_{ini} = 2$.

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



Utilizziamo il metodo del gradiente:

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

A.A. 2010-2011

57/62

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



Minimizzazione - underdamping



Consideriamo $\alpha = 1$

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

Utilizziamo il metodo del gradiente:

Passo 1:

Calcoliamo l'incremento da dare al parametro a :

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

Passo 2:

Calcoliamo l'incremento da dare al parametro a :

$$da = -[-2(3 - 4 \cdot 1) \cdot 1] = -[-6 + 8] = -2 \quad a^{21} = 4 - 2 = 2$$

Oscillazioni!!!

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

A.A. 2010-2011

58/62

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



Minimizzazione -2 passi



Consideriamo $\alpha = 0.4$

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

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



$\min(f(x, w))$ 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, $f(\cdot)$.

Estensione della tecnica del gradiente a più variabili.

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

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 α ottimale, per cui la funzione è decrescente (line search).



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.