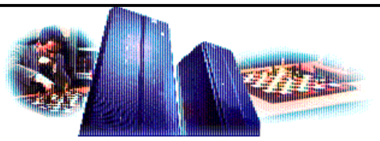
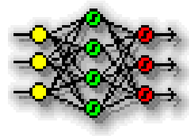




Sistemi Intelligenti
Stimatori e identificazione - II

Alberto Borghese

Università degli Studi di Milano
Laboratory of Applied Intelligent Systems (AIS-Lab)
Dipartimento di Informatica
borghese@di.unimi.it

Overview

Modelli

Sistemi lineari

Densità di probabilità

Massima versosimiglianza

Modello (predittivo)

$z = f(u \mid w)$

ν - noise

u

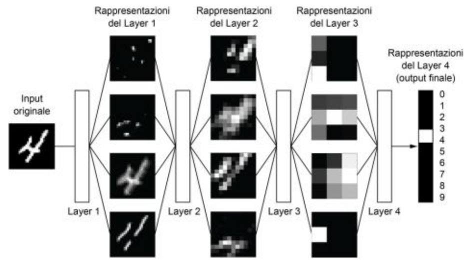
z

z_m

u – causa => z - effetto; z_m – effetto (misurato con errore)

Regressione predittiva

Classificazione



Modello (predittivo)

$z = f(u \mid w)$

ν - noise

u

z

z_m

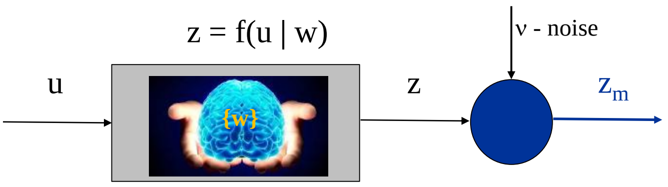
u – causa => z (effetto); z_m – effetto (misurato con errore)

FASI CREAZIONE MODELLO PREDITTIVO

Definizione progetto → Preparazione dei dati → Costruzione del modello → Validazione del modello → Utilizzo

Realizzazione del modello | utilizzo

I 3 problemi associati ai Modelli



u – causa => z (effetto); z_m – effetto (misurato con errore)

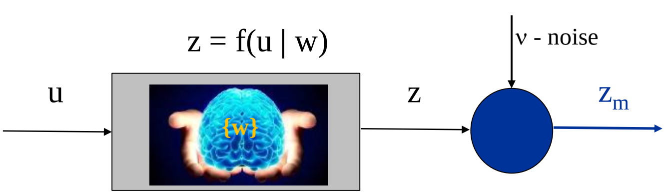
Control / Classification / Prediction: determine {z} from {u},{w} – utilizzo forward

Inverse problem: determine cause {u} from {z_m},{w} – utilizzo backwards

Inverse problem: Identification: determine {w} from {u},{z_m} - *Learning*

Ruolo dei modelli

- **Identificazione (learning):** stimo i parametri di un modello a partire dai dati: identifico il modello.
- **Utilizzo 1 (backwards):** utilizzo il modello per inferire informazioni sulla causa di un effetto misurato.
- **Utilizzo 2 (forward):** utilizzo il modello per inferire informazioni su nuovi dati (controllo, regressione predittiva, classificazione).

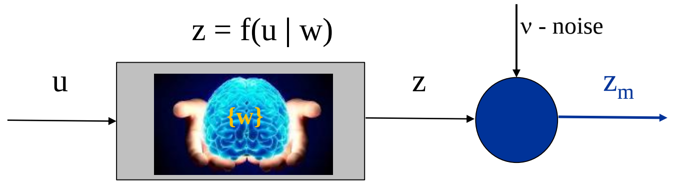


Osservazioni

- Un modello puo' essere utilizzato forward: fornisco un input e calcolo l'output associato a quell'input: output ideale, z.

- Il modello puo' essere utilizzato backwards: misuro un insieme di uscite, z_m e da qui:
  - A) conosco l'insieme di input corrispondenti, {u}, calcolo i parametri w.
  - B) conosco i parametri del modello, calcolo i valori di ingresso {u} corrispondenti alle uscite.

In questo senso input e parametri sono duali tra loro.



Overview

Modelli

Sistemi lineari

Densità di probabilità

Massima versosimiglianza

Esempio di modello lineare generale (sensore di misura – video-camera)

- $u = \{u_1, u_2, \ldots, u_M\}$, $u_k \in R^M$ e.g. Pixels true luminance
- $z_n = \{z_{n1}, z_{n2}, \ldots, z_{nM}\}$ $z_{nk} \in R^M$ e.g. Pixels measured luminance (noisy)

- z_n = **A** u + n + h-> determining x is a **deblurring problem** (the measuring device introduces in each pixel measurment error and some blurring – contribution of neighbour pixels)
- **This is the very general equation that describes any sensor.**

Role of A:

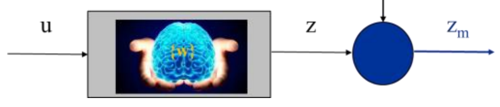
- Matrix that produces the ideal output z_i as a linear combination of values of u (blurring is a form of a weighted average of pixels inside a neighbourhood).

Role of h: offset: background radiation (dark currents)- is compensated by calibration, regulation of the zero point.

Role of n: measurement noise.

$z = Au + \cancel{h}$

- $z_n = A\,u + \cancel{h} + n$ after calibration



Sistema lineare

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

$\{a_{ij}\}$ – coefficienti M x N
$\{x_j\}$ – incognite, N x 1
$\{b_j\}$ – termini noti, M x 1

A x = b

M x N (Matrice di disegno)

Vettore delle incognite

M x 1 Vettore dei termini noti

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

Esempio:

$3x_1 + 2\,x_2 + \ldots\ldots\ 4x_N = 5$
$4x_1 -2\,x_2 + \ldots\ldots\ 0.5\,x_N = 3$
………………
$2x_1 + 3\,x_2 + \ldots\ldots\ -3\,x_N = -1$

Sistema lineare e modelli: forward

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

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

....

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

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

$\{x_j\}$ – incognite, N

$\{b_j\}$ – termini noti, M

Utilizzo forward:

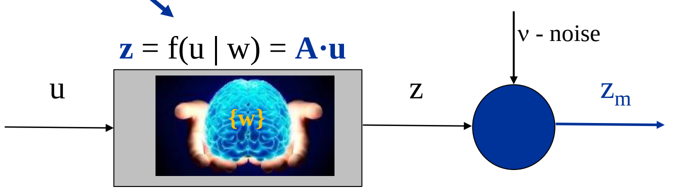
$\{a_{ij}\}$ – parametri $\{w\}$

$\{x_j\}$ – input, $\{u\}$

$\{b_j\}$ – uscite, $\{z\}$

Modello lineare

$\mathbf{z} = f(u \mid w) = \mathbf{A \cdot u}$



Sistema lineare e modelli

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

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

....

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

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

$\{x_j\}$ – incognite, N

$\{b_j\}$ – termini noti, M

Utilizzo backwards (identificazione del modello)

$\{a_{ij}\}$ – input $\{u\}$

$\{x_j\}$ – parametri, $\{w\}$

$\{b_j\}$ – uscite misurate $\{z_m\}$

Utilizzo backwards (determinazione della causa)

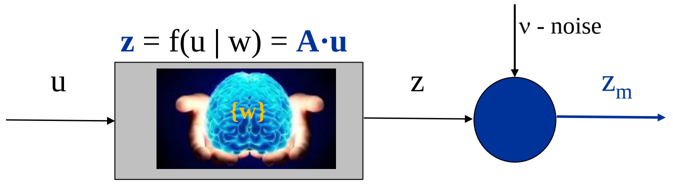
$\{a_{ij}\}$ – parametri $\{w\}$

$\{x_j\}$ – input, $\{u\}$

$\{b_j\}$ – uscite misurate $\{z_m\}$

Modello lineare

$\mathbf{z} = f(u \mid w) = \mathbf{A \cdot u}$



Matrici

Insieme di valori organizzati per righe e colonne

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

$\alpha A = [\alpha a_{i,j}]$ $\quad \alpha$ = cost $\quad C = A + B = [a_{i,j} + b_{i,j}]$

$C = AB = [c_{i,j}]$ 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)

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

La somma gode della prorietà associativa e commutativa.

Il prodotto è associativo rispetto alla somma ma non gode della proprietà commutativa:

(A+B)C = AC + BC.

AB $\neq$ BA

Altre proprietà delle matrici

Una matrice W = $\{w_{ij}\}$ si dice diagonale se $w_{ij} = \begin{cases} w_{ii} \text{ per } i = j \\ 0 \text{ altrimenti} \end{cases}$

Matrice identità. $I = \{a_{ij}\} : \begin{cases} 1 \text{ per } i = j \\ 0 \text{ altrimenti} \end{cases}$ $\qquad A \cdot I = I \cdot A = A$

Matrice trasposta e proprietà:

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

Minore complementare

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

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

Somma dei prodotti degli elementi di una riga o colonna per il loro complemento algebrico (formula di Leibniz).
Il complemento algebrico è il minore complementare di un elemento moltiplicato per -1 elevato alla somma del numero di riga con il numero di colonna

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

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \longleftarrow$$ 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) $[(1*1)-(3*1)] = -16 + 2 = -14$

det(ABC) = det(A) det(B) det(C)

Matrice inversa

Viene definita solo per matrici **quadrate** (N x N):

$A^{-1}A = I$

Esiste ed è unica se $\det(A) \neq 0$.

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

$(A\, B\, C)^{-1} = C^{-1}\, B^{-1}\, A^{-1}$

Risoluzione di un sistema 2x2

$$\begin{cases} a_{11}x_1+a_{12}x_2 = b_1 \\ a_{21}x_1+a_{22}x_2 = b_2 \end{cases}$$

$\mathbf{y = Ax}$

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

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

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

Rango di una matrice

Data una matrice A di ordine n (n x n),

una matrice A n x n ha rango m < n se e solo se esiste un suo minore di ordine m non nullo (determinante $\neq 0$) mentre sono nulli tutti i minori di ordine m + 1.

Una matrice A n x n ha rango n (rango pieno) se e solo se il suo determinante è diverso da 0

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

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

Ortonormalità

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 (rotazione di sistema di riferimento)**.

Sistema lineare: soluzione robusta (SVD)

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

$\mathbf{A = U^T\, W\, V}$

Ortonormale M x N

Diagonale (N x N)

Ortonormale N x N

Se N = M

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

$\mathbf{A^{-1} = (U^T\, W\, V)^{-1} = V^T\, W^{-1}\, U}$

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

w_{ii} sono detti valori singolari.

$\mathbf{Ax = b}$ ➔ $\mathbf{x = A^{-1}\, b = V^T\, W^{-1}\, U^T\, b}$

Condizionamento di una matrice

La matrice inversa esiste ed è unica se det(A) ≠ 0.

$$A^{-1} = (U^T \, W \, V)^{-1} = V^T \, W^{-1} \, U$$

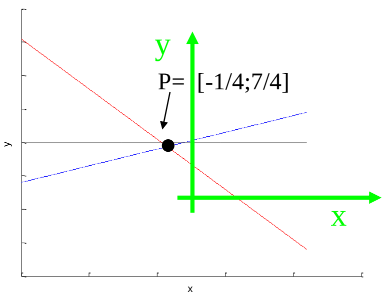
$$\det(A) = \det(U^T)\det(W)\det(V) = 1 \cdot \left(\prod_{i=1}^{N} w_{ii}\right) \cdot 1$$

Numero di condizionamento di una matrice: rapporto tra il valore singolare maggiore e minore (w_{11} / w_{nn}) - cf. Funzione cond in Matlab).

E' una misura di sensibilità della soluzione di un sistema lineare a variazioni nei dati.

Esempio di soluzione di un sistema lineare

$$\begin{cases} y = x + 2 \\ y = -3x + 1 \end{cases}$$



Risolvo per sostituzione: x = -2 + 1 y.

-3(-2 + y) - y = -1 ➔ y = 7/4
x - 1/4 = 2 ➔ x = -1/4

Ottengo 1 soluzione

Rette e sistemi lineari

$$\begin{cases} y = x + 2 \\ y = -3x + 1 \end{cases}$$

$y = x_2$
$x = x_1$

Scrivo il sistema lineare: $Ax = b$

$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{cases} 1\,x_1 - 1\,x_2 = -2 \\ -3\,x_1 - 1\,x_2 = -1 \end{cases}$$

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

$X = [x_1\ x_2]$ è il punto all'intersezione delle due rette

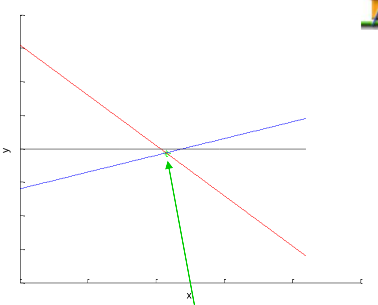
Esempio

$$\begin{cases} y = x + 2 \\ y = -3x + 1 \end{cases}$$

$x_1 = x$
$x_2 = y$

$$\begin{cases} 1\,x_1 - 1\,x_2 = -2 \\ -3\,x_1 - 1\,x_2 = -1 \end{cases}$$

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



$\det(A) = 1(-1)-(-1)(-3)=-1-3=-4 \neq 0$

$\boxed{\text{Rango di A è pieno}}$

P = [-1/4 7/4]

$$x = A^{-1}\, b = -\frac{1}{4}\begin{bmatrix} -1 & +1 \\ +3 & +1 \end{bmatrix} \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

$x_1 = -1/4$
$x_2 = 7/4$

Esempio di soluzione non univoca (det(A) = 0)

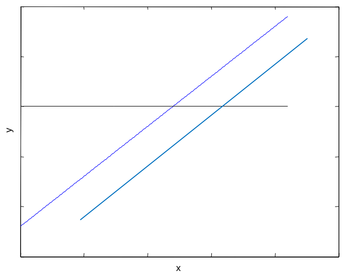
$$\begin{cases} y = x + 2 \\ 2y = 2x + 3 \end{cases}$$

$x_1 = x$

$x_2 = y$

$$\begin{cases} 1\,x_1 - 1\,x_2 = -2 \\ 2\,x_1 - 2\,x_2 = -3 \end{cases}$$

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



det(A) = 1(-2)-(-1)(2)=-2+2=0 La soluzione non esiste o ∞ soluzioni.

$$\begin{cases} y = x + 2 \\ 2y = 2x + 4 \end{cases}$$

La soluzione, non è unica: tutti i punti della retta soddisfano contemporaneamente le 2 equazioni. In questo caso ∞ soluzioni: rette sovrapposte.

Sistema M x N, M > N

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

Esempio:

$$\begin{cases} 3x_1 + 2\,x_2 + \ldots\ldots\ \ 4x_N = 5 \\ 4x_1 - 2\,x_2 + \ldots\ldots\ \ 0.5\,x_N = 3 \\ \ldots\ldots\ldots \\ 2x_1 + 3\,x_2 + \ldots\ldots\ \ -3\,x_N = -1 \end{cases}$$

A x = b

A è M x N, M > N, non è una matrice quadrata.

N equazioni sono sufficienti per determinare la soluzione.

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

1, nessuna, ∞ soluzioni.

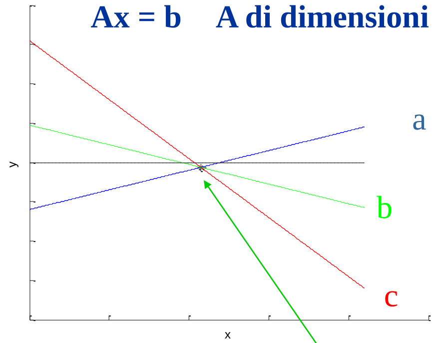
Posso sempre calcolare la soluzione in forma matriciale.

Sistemi lineari con m > n

A è rettangolare: numero di righe maggiore del numero di colonne

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

Ax = b   A di dimensioni m x n



P = [-1/4 7/4]

Una delle 3 righe di A è combinazione lineare delle altre. Risolvo per sostituzione

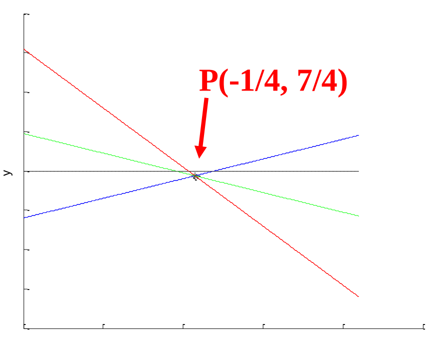
$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -1 \\ -1.5 \end{bmatrix}$$

Esiste un'equazione "di troppo"

Nessuna, 1 o ∞ soluzioni

Rango di A è pieno -> 1 soluzione

Relazione tra le equazioni (combinazione lineare)



α_1 (y - x - 2) +
α_2 (y +3x – 1) =
(y + x - 3/2)

In questo caso:
α_1 = -1/2
α_2 = -1/2

Tutte le rette per la soluzione P possono essere descritte come un fascio (di rette).

Un fascio di rette è univocamente identificato da due rette (che si incontrino in un punto).

La terza equazione è combinazione lineare delle prime due.

Sistema lineare: soluzione algebrica

Caso generale:

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

$$\Downarrow$$

$$\mathbf{Ix = (A^T A)^{-1} A^T b}$$

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

Quale criterio viene soddisfatto da **x**?

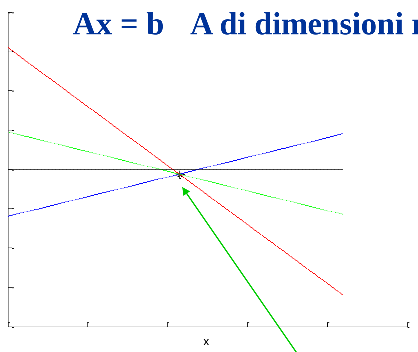
$(A^T A)^{-1} A^T$ è la matrice pseudoinversa

Sistemi lineari con m > n

$$\begin{cases} y = x + 2 \\ y = -3x + 1 \\ y = -x + 3/2 \end{cases} \qquad \begin{cases} x_1 - x_2 = -2 \\ -3x_1 - x_2 = -1 \\ -x_1 - x_2 = -3/2 \end{cases}$$

Ax = b   A di dimensioni m x n

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



P = [-1/4 7/4]

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

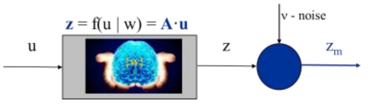
$$C = (A^T A)^{-1} = \begin{bmatrix} 0.1250 & -0.1250 \\ -0.1250 & 0.4583 \end{bmatrix} \qquad x = C * A^T * b$$

P = [-0.25 +1.75]

intersezione

Sistema lineare con errore sul termine b_i

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \ldots\ldots \quad a_{1N}x_N \neq b_{m,1} \\
a_{21}x_1 + a_{22}x_2 + \ldots\ldots \quad a_{2N}x_N \neq b_{m,2} \\
\ldots\ldots\ldots\ldots \\
a_{M1}x_1 + a_{M2}x_2 + \ldots\ldots \quad a_{MN}x_N \neq b_{m,M}
\end{cases}
$$



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

M x 1

Vettore dei termini noti

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

Esiste una soluzione? Qual è il valore di x che posso calcolare?

Riformulazione del problema con errore

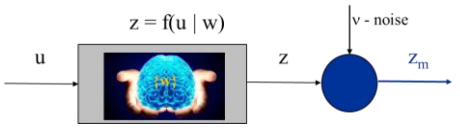
Misura reale

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \ldots\ldots \quad a_{1N}x_N = b_{m,1} + \nu_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots\ldots \quad a_{2N}x_N = b_{m,2} + \nu_2 \\
\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots \\
a_{M1}x_1 + a_{M2}x_2 + \ldots\ldots \quad a_{MN}x_N = b_{m,M} + \nu_M
\end{cases}
$$

Modello

Misura reale

Errore di misura



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

M x 1

Vettore dei termini noti

M x N

(Matrice di disegno)

N x 1

Vettore delle incognite

Esiste una soluzione? Qual è il valore di x che posso calcolare?

Soluzione come problema di ottimizzazione

A x = b + N

Funzione costo: $|N|^2 = \sum_k \nu_k^{\ 2} = ||Ax - b||^2$

Assegno un costo al fatto che la soluzione x, non soddisfi tutte le equazioni, la somma dei residui associati ad ogni equazione (misura) viene minimizzata.

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

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

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

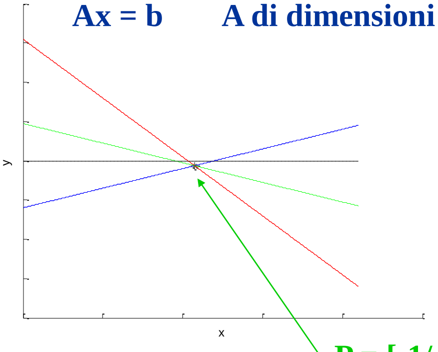
$\mathbf{x = (A^TA)^{-1}A^Tb}$

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

Sistemi lineari con m > n

$$\begin{cases} x_1 - x_2 = -2 \\ -3x_1 - x_2 = -1 \\ -x_1 - x_2 = -3/2 \end{cases}$$

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



P = [-1/4 7/4]

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

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

$$C = (A^TA)^{-1} = \begin{bmatrix} 0.1250 & -0.1250 \\ -0.1250 & 0.4583 \end{bmatrix} \qquad x = C * A^T * b$$

P = [-0.25 +1.75]

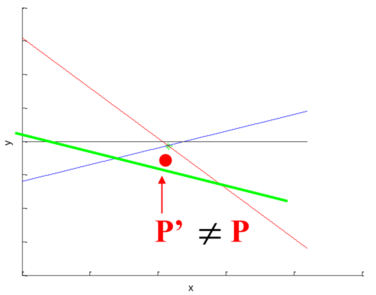
intersezione

$||Ax - b|| = 0$

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

$$\begin{cases} x_1 - x_2 = -2 + \mathbf{0} \\ -3x_1 - x_2 = -1 + \mathbf{0.5} = \mathbf{-0.5} \\ -x_1 - x_2 = -3/2 + \mathbf{0} = \mathbf{-3/2} \end{cases}$$

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



$$A = \begin{bmatrix} 1 & -1 \\ -3 & -1 \\ -1 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} -2 \\ -0.5 \\ -1.5 \end{bmatrix}$$

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

$$\sum_k \nu_k^{\,2} = \|Ax - b\|^2 = 0.04116666$$

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

$x = C * A^T * b$ **P' = [-0.375 +1.70833]**

No intersezione

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

(P' = [-0.25 +1.75])

Soluzione mediante pseudo-inversa

$$\sum_k \nu_k^{\,2} = \|Ax - b\|^2 = \sum_k \| A_{k,*}x - b_k \|^2 =$$

Sommo per tutte le righe

$$[(A_{11}x_1 + A_{12}x_2) - b_1]^2 + [(A_{21}x_1 + A_{22}x_2) - b_2]^2 + [(A_{31}x_1 + A_{32}x_2) - b_3]^2$$

P' = [-0.375 +1.70833]

No intersezione

$x_1 - x_2 = -2$ $-0.375 - 1.70833 + 2 = \nu_1 = -0.083333$

$-3x_1 - x_2 = -1/2$ $+1.125 - 1.70833 + 0.5 = \nu_2 = 0.083333$

$-x_1 - x_2 = -3/2$ $+0.375 - 1.70833 + 1.5 = \nu_3 = 0.16666$

$$\sum_k \nu_k^{\,2} = \|Ax - b\|^2 = \sum_k \| A_{k,*}x - b_k \|^2 = 0.04116666$$

Lo scarto misura la somma quadratica delle distanze (verticali, lungo z, =b nel sistema lineare, la misura) tra il punto che rappresenta la soluzione e le 3 rette.

Sistema lineare: soluzione robusta per m > n

$\mathbf{A\,x = b} \Longrightarrow \mathbf{A^T A\,x = A^T b} \Longrightarrow \mathbf{x = (A^T A)^{-1} A^T b}$

Rango (A^TA) = Rango(A)

Numero di condizionamento varia circa con la norma di (A^T*A).

Soluzione tramite Singular Value Decomposition

Numero di condizionamento varia circa con A.

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

$\mathbf{U\,W\,V\,x = b}$

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

Ortonormale M x N (rettangolare)

Diagonale (N x N)

Ortonormale N x N

- La matrice C = $(A^TA)^{-1}$ non viene formata.
- W^{-1} contiene i reciproci degli elementi di W.

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

Overview

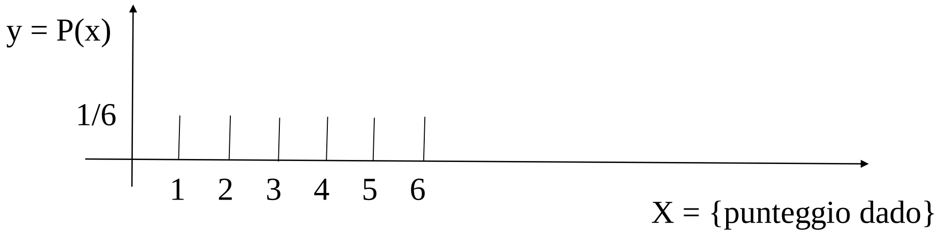
Modelli

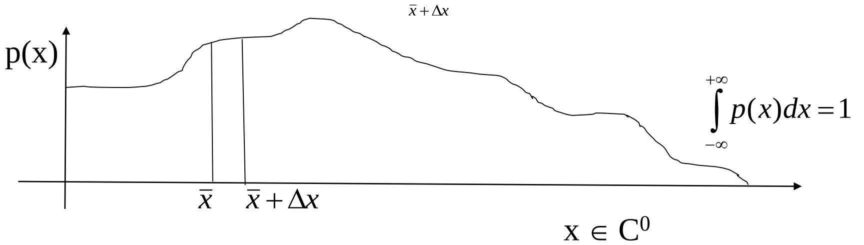
Sistemi lineari

Distribuzione di probabilità

Massima verosimiglianza

La probabilità nel caso continuo





$$P(x \in [\bar{x}, \bar{x}+\Delta x]) \int_{\bar{x}}^{\bar{x}+\Delta x} p(x)dx$$

Definizione di p(x)

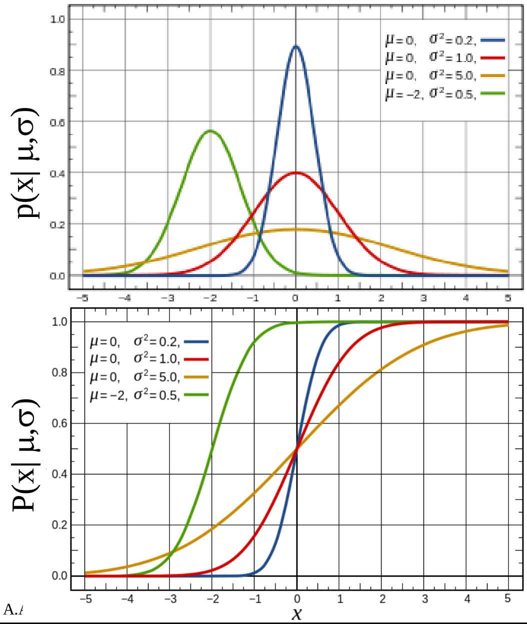
Caso discreto: prescrizione della probabilità per ognuno dei finiti valori che la variabile X può assumere: P(X).

Caso continuo: i valori che X può assumere sono infiniti. Devo trovare un modo per definirne la probabilità. Descrizione **analitica** mediante la funzione densità di probabilità. Si considera la probabilità che x cada in un certo intervallo.

Valgono le stesse relazioni del caso discreto, dove alla somma si sostituisce l'integrale.

$$P(X = x \in [\bar{x}, \bar{x}+\Delta x]) \int_{\bar{x}}^{\bar{x}+\Delta x} \int_{-\infty}^{+\infty} p(x, y)dxdy$$

Distribuzioni notevoli: la Gaussiana



$$p(x\,|\,\mu,\sigma)=\frac{1}{\left(\sqrt{2\pi}\right)\Sigma^{D}}\cdot\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\Sigma}\right)^{2}\right]$$

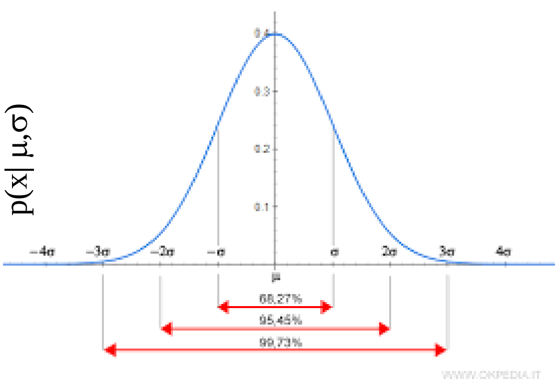
$$\int p(x|\mu,\sigma)=1$$

D = dimensione, in questo caso D = 1

Il valore con probabilità più elevato è intorno al valore medio μ.

Carl Friedrich Gauss.

Concentrazione dei valori



Suppporto infinito

$$p(x\,|\,\mu,\sigma)=\frac{1}{\left(\sqrt{2\pi}\right)\Sigma^{D}}\cdot\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\Sigma}\right)^{2}\right]$$

D = dimensione, in questo caso D = 1

Pr(| X-μ | < σ) = 0.68268

Pr(| X-μ | < 2σ) = 0.95452

Pr(| X-μ | < 3σ) = 0.9973

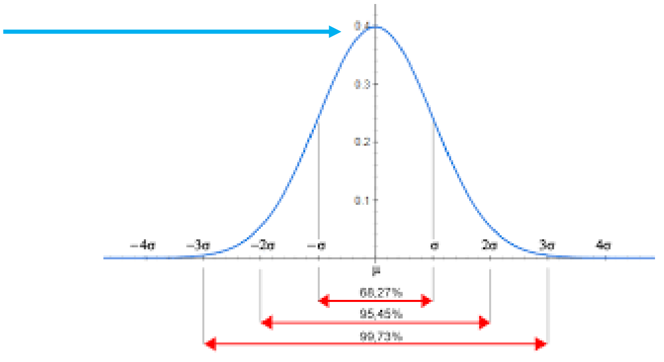
I momenti di una variabile statistica

$$\mu^k(X) = \int_{-\infty}^{+\infty} (x-a)^k \, p(x) dx$$

Momento rispetto ad a, solitamente alla media

Valore atteso (Expected value) di X = media distribuzione

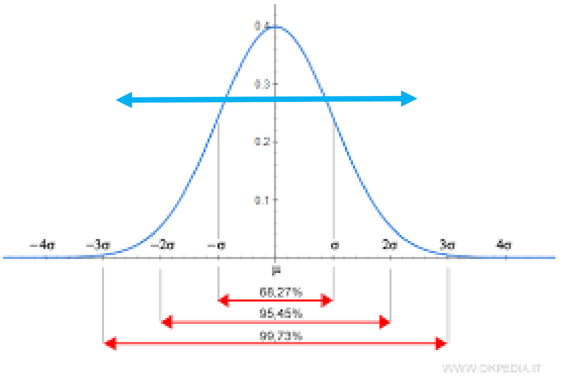
$$E[X] = \int_{-\infty}^{+\infty} x p(x) dx$$



I momenti di una variabile statistica

$$E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 \, p(x)$$

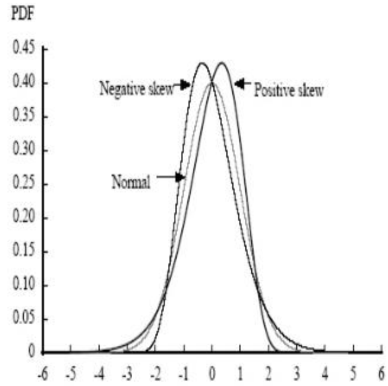
Varianza - σ^2



I momenti di una variabile statistica

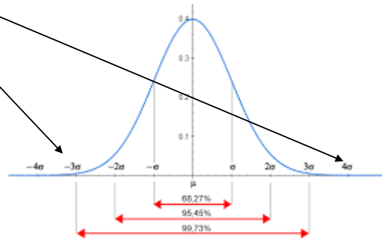
Asimmetria

$$E\left[(X-\mu)^3\right]=\int_{-\infty}^{+\infty}(x-\mu)^3\,p(x)$$

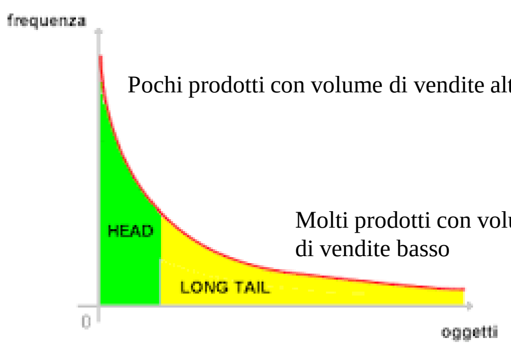


Kurtosi – peso delle code di p(x)

$$E\left[(X-\mu)^4\right]=\int_{-\infty}^{+\infty}(x-\mu)^4\,p(x)$$

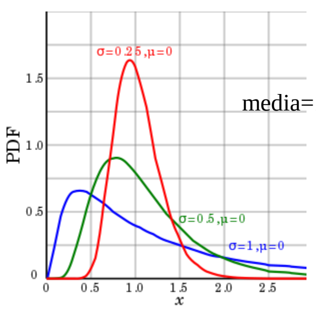


Distribuzioni notevoli::a coda lunga



Pochi prodotti con volume di vendite alto

Molti prodotti con volume di vendite basso



media=1

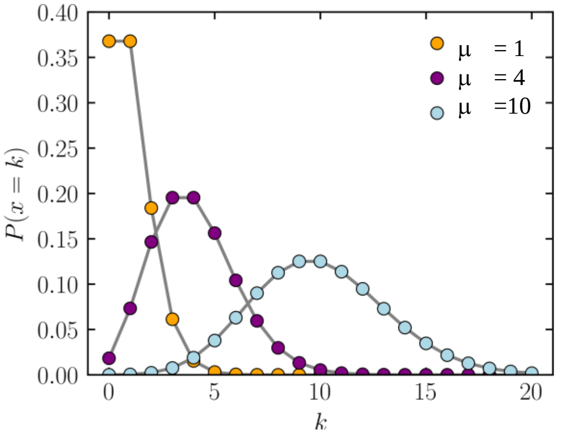
Distribuzione log-normale
(il suo logaritmo è distribuito come una normale)

t = posizione (traslazione)
μ = log(media)
σ = log(varianza)

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

E.g. marketing (Amazon, Netflix): strategia di vendita al dettaglio che preferisce vendere un gran numero di oggetti unici in quantità relativamente piccole di ogni oggetto venduto (long selling).

Distribuzioni notevoli::Poisson



Distribuzione di eventi rari, quando il valor medio, μ. cresce (≈30) viene assimilata a una Gaussiana.
La varianza è uguale alla media.
Deriva dalla distribuzione binomiale (probabilità che verifichi un evento)

P(n | μ) = $\frac{\mu^n}{n!} e^{-\mu}$

$\sigma^2 = \mu$

Overview

Modelli

Sistemi lineari

Distribuzione di probabilità

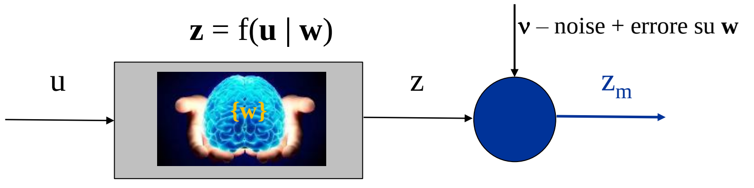
Massima verosimiglianza

Probabilità di un certo insieme di misure

$z = f(u \mid w)$ misuro $\{z_i\}$ in corrispondenza di $\{u_i\}$.

$\{z_i\}$ è ottenuto come uscita del modello, tramite i parametri $\{w_j\}$

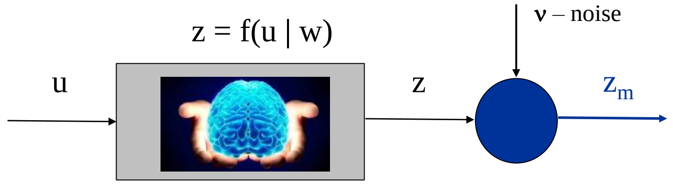
Avrò che: $f(u_i , w) = z_i = z_{i,m} - \nu_i$



Se le misure sono indipendenti posso scrivere che la probabilità di ottenere l'insieme di misure indipendenti tra loro (tutte dipendono da x): z_{1m}, z_{2m}, z_{3m} ….. è:

$p(z_{1m}, z_{2m}, z_{3m}) = \prod_i p\left(z_{im}\right)$ (cf. dadi nel caso discreto)

Dipendenza delle misure



Le misure dipendono dal valore delle variabili in ingresso {u} e dai parametri del modello {w}. Supponiamo {u} e {w} deterministici.

$$p(z_{1m}, z_{2m}, z_{3m}) = \prod_i p(z_{im} \mid z) = \prod_i p(z_{im} \mid u_i, w) =$$

Errore di misura

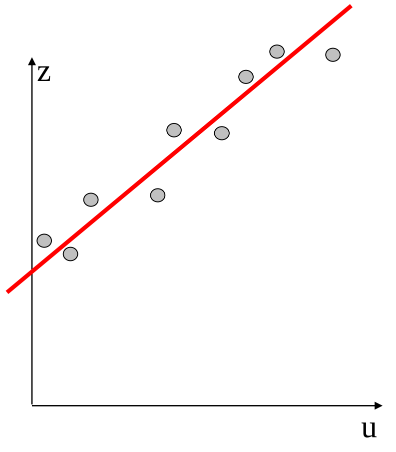
Scrivo la probabilità esplicitamente come condizionata al valore di u e di w.

Tanto più i parametri saranno corretti tanto maggiore sarà la correttezza di z in uscita dal modello.

Esempio: fitting di una retta

Vogliamo stimare i parametri di una retta: $z = f(u \mid w) =$ **m** $u +$ **q**, con m e q incogniti: $W = \{m, q\}$

La retta è un modello lineare.



Supponiamo che le z_i siano affette da **errore Gaussiano a media nulla.**

Abbiamo a disposizione N misure rumorose effettuate $\{z_{im}\}$. Errore sull'asse delle z.

Possiamo anche scrivere che: $z_{im} = z_i + G(\mu, \sigma^2)$ indica una distribuzione monodimensionale Gaussiana a media μ e varianza σ^2. Errore di misura a media nulla: $G(0, \sigma^2)$

$z_{im} = z_i + \nu_i = (m\, u_i + q) + \nu_i$ dove ν_i è l'errore di misura, **Gaussiano a media nulla.**

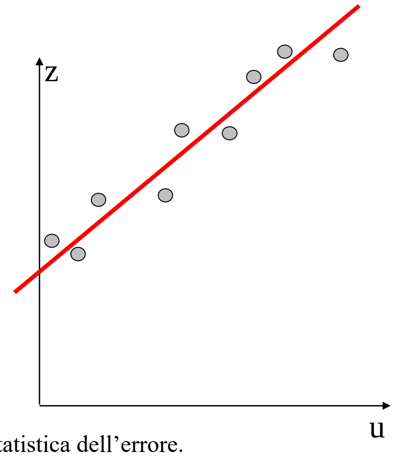
Stima dei parametri del modello

Per ogni punto, dovrebbe valere $z_i = m\, u_i + q$.

Ma c'è l'errore di misura, misuriamo in realtà $z_i + \nu_i = z_{im}$

$\boldsymbol{\nu_i = z_{im} - (m\, u_i + q)}$

misura *uscita del modello*



I ν_i si distribuiscono secondo la statistica dell'errore.

Funzione di verosimiglianza

- Siano date **N variabili casuali <u>indipendenti</u>**… Quale è la **probabilità di misurare il vettore [z_{1m}, …, z_{Nm}]**?

$$p(z_{1m}, z_{2m}, ..., z_{Nm}) = p(z_{1m}) \cdot p(z_{2m}) \cdot ... p(z_{Nm}) = L(z_{1m}, z_{2m}, ..., z_{Nm})$$

- La probabilità congiunta è il prodotto delle probabilità semplici (***misure indipendenti tra loro***).

- Questa è la **Funzione di verosimiglianza** o **funzione di likelihood**, **L(.)**

Funzione di verosimiglianza e modello

- In questo caso le z sono legate alla variabile indipendente u da f(u,w).
  Nel caso della retta f(.) = mu + q

- Troviamo i parametri {w} tali per cui è massima la probabilità di misurare il vettore di misure:
  $\mathbf{z_m} = \{z_{im}, i=1…N\}$.

$$L(z_{1m}, z_{2m}, z_{3m}, \dots, z_{Nm} \mid (w; u_1, u_2, u_3, \dots u_{Nm})) =$$

$$= p(z_{1m} \mid w; u_1)\, p(z_{2m} \mid w; u_2)\, p(z_{3m} \mid w; u_3) \dots.. p(z_{Nm} \mid w; u_N)$$

- $L = L(\mathbf{z_m} \mid u, w)$. dati u e w, ottengo un certo valore di probabilità per z_m.

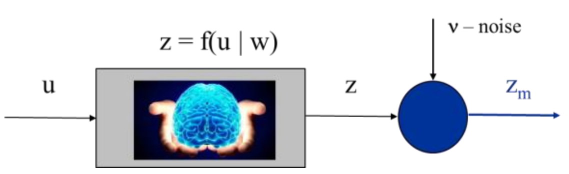
Osservazioni

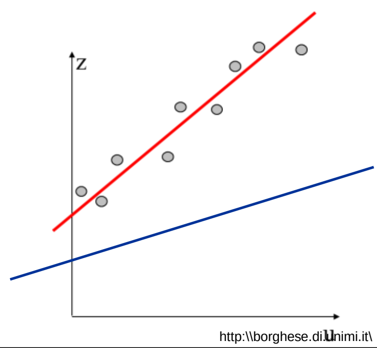
$$L(z_{1m}, z_{2m}, z_{3m}, \dots , z_{Nm} \mid (w; u_1, u_2, u_3,\dots u_{Nm})) =$$
$$= p(z_{1m} \mid w; u_1)\, p(z_{2m} \mid w; u_2)\, p(z_{3m} \mid w; u_3) \dots\dots p(z_{Nm} \mid w; u_N)$$

- Più in generale, le variabili possono avere un residuo, v, descritto da densità di probabilità diverse.
- La relazione tra ingresso e uscita è la stessa per tutte le misure, ed è rappresentata dal modello.
- La forma del modello dipende da un insieme di parametri, w.

Massima verosimiglianza - retta

- La funzione di verosimiglianza dipende da u e w.
- Modificando il valore di w adatto la funzione f(.), la retta, in modo che gli {z} (misurati sulla retta rossa) in corrispondenza degli input {u} siano i più vicina a z_m.
- Massimizzo la verosimiglianza di ottenere gli $\{z_m\}$.
- Nel caso della retta ruoto e traslo la retta in modo tale che si avvicini "il più possibile" ai punti misurati.
- La retta rossa rende più verosimile che gli $\{z_m\}$ vengano misurati della retta blu.





Stima a massima verosimiglianza

$$p(z_{1m}, z_{2m}, ..., z_{Nm}) = p(z_{1m}) \cdot p(z_{2m}) \cdot ... p(z_{Nm}) = L(z_{1m}, z_{2m}, ..., z_{Nm}) = \prod_i p(z_{im})$$

$$\mathrm{Max}_{\{w\}} \; (L(z_{1m}, z_{2m}, z_{3m}, \dots , z_{Nm} \mid (w; u_1, u_2, u_3, \dots u_{Nm}))) =$$
$$\mathrm{Max}_{\{w\}} = p(z_{1m} \mid w; u_1)\, p(z_{2m} \mid w; u_2)\, p(z_{3m} \mid w; u_3) \,\dots\dots\, p(z_{Nm} \mid w; u_N)$$

Devo trovare un modo efficiente per massimizzare la funzione di verosimiglianza, o likelihood, L(.) rispetto ai parametri {w} che determinano la forma del modello.

Cosa vuol dire che sono più verosimili? Quanto sono più verosimili?

Massimizzando la funzione di verosimiglianza rispetto a tali parametri se ne effettua la stima in modo tale che il vettore osservato $z_m = \{z_{im}\}$ i =1…N sia massimamente probabile (massima verosimiglianza) ovverosia i valori prodotti dal modello siano il più vicino possibile ai valori misurati.

Stima alla massima verosimiglianza per modello lineare

- Equazione di una retta: z = mu + q
- Scriviamo prima di tutto la densità di probabilità di ottenere z_{im} per ciascun dato, per errore, ν_i, Gaussiano a media nulla:

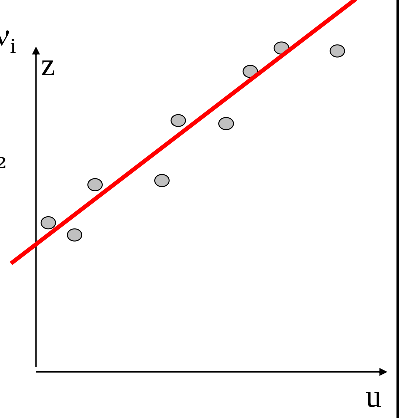
$z_{im} = z_i + \nu_i$ <== $z_{im} = (m\, u_i + q) + \nu_i$

z_i

$$(z_{im} \mid m, q; u_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{z_{im}-(m\, u_i+q)}{\sigma}\right)^2}$$

dove m e q non sono note.

ν_i



Overview

Modelli

Sistemi lineari

Distribuzione di probabilità

Massima verosimiglianza