Lezione 1

mercoledì 7 ottobre 2020 09:38

INTRODUZIONE

1) IDENTIFICAZIONE DEI MODELLI

MODELLO: Descrizione matematica di un fenomeno o di un sistema
- economico
- sociale
- fisico: $I(t) = C \cdot \frac{dV(t)}{dt}$ — 22 pF

SISTEMA: meccanismo adatto da trasformare input in output

input → causa; output → effetto

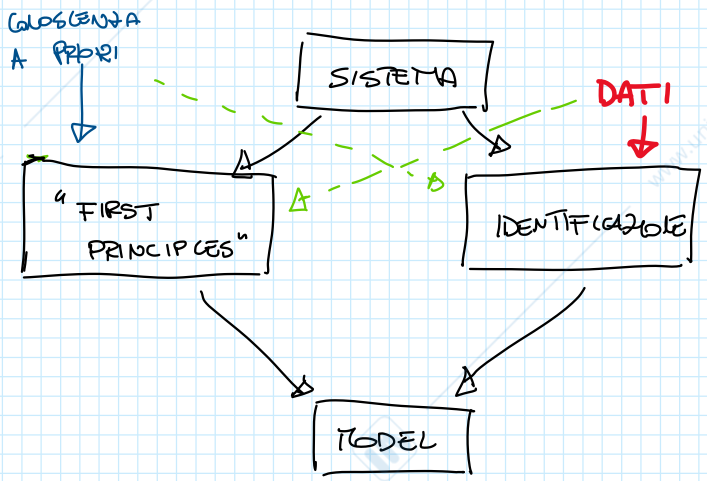
$u \rightarrow \boxed{S} \rightarrow y$

Due approcci:

a) WHITE-BOX - Leggi del sistema — GENERALITÀ +, COMPLESSITÀ −

b) BLACK-BOX - approccio basato sui dati — VASTA APPLICABILITÀ +, DIPENDENZA DAI DATI ACQUISITI −

$u \rightarrow \boxed{f(\cdot)} \rightarrow \hat{y} \qquad (y - \hat{y})^2$



$$G(s) = \frac{Y(s)}{U(s)} = \frac{\mu(1+sT)}{1+s\tau}$$

IDENTIFICAZIONE DEI MODELLI

⇓

PROBLEMA DI STIMA

2) ANALISI DEI DATI

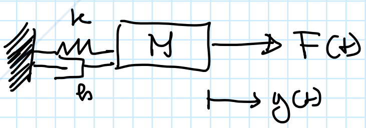
- Determinare le caratteristiche statistiche dei dati

- Determinare le caratteristiche statistiche dei dati
  - L RUMORE, INCERTEZZA

- Individuare una STRUTTURA all'interno dei dati

MODELLI STATICI: la sola conoscenza delle variabili di ingresso u è sufficiente a determinare l'uscita y $\quad V(t) = R \cdot I(t)$

MODELLI DINAMICI: ho bisogno di conoscere anche la condizione iniziale



$$M\ddot{y}(t) = F(t) - k\,y(t) - b\,\dot{y}(t)$$

$$y(0),\ \dot{y}(0)$$

$$\hat{G}(s)$$

$$G(s) = \frac{3}{1+7s}$$

RICHIAMI DI STATISTICA $\boxed{1}$

$v(s)$ ⇐ ESITO s di un esperimento casuale

$$v(s) = \begin{cases} 1 & s = \text{TESTA} \\ 0 & s = \text{CROCE} \end{cases}$$

$\bar{s}$ $\quad v(\bar{s})$: valore assunto da $v(s)$ a fronte del particolare esito $s = \bar{s}$

- v assume valori DISCRETI (V.C. DISCRETA)
  - L funzione di probabilità di massa (pmf) $\quad p(x) = P(v = x)$

x_i : valori di v $\quad x_1 = 1,\ x_2 = 0 \Rightarrow \sum_{i=1}^{m} p(x_i) = 1$

Esempio: LANCIO DADO

$x_1 = 1 \quad p(x_1) = P(v = x_1) = 1/6$

$x_2 = 2$

$\vdots$

$x_6 = 6 \quad p(x_6) = P(v = x_6) = 1/6$

$$\sum_{i=1}^{6} p(x_i) = 6 \cdot \frac{1}{6} = 1$$

- v assume valori CONTINUI (V.C. CONTINUA)

- v assume valori CONTINUI (V.C. CONTINUA)

Esempio: v descrive l'altezza di una persona in cm

NON HA SENSO DEFINIRLO $\Rightarrow \nexists P(v = 173,4678134 \text{ cm}) \to 0$ $\qquad \frac{1}{\infty} \to 0$

L funzione di densità di probabilità $f(x)$ (pdf)

$$P(v \in [a,b]) = \int_a^b f(x)\, dx$$

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

L funzione di densità cumulata o distribuzione di probabilità (cdf)

$$F(z) = \int_{-\infty}^{z} f(x)\, dx = P(v \leq z)$$

- VALORE ATTESO di una v.c. continua è:

$$\mathbb{E}[v] = \int_{-\infty}^{+\infty} \boxed{x} \cdot f(x)\, dx$$

LINEARITÀ: $\mathbb{E}[\alpha v_1 + \beta v_2 + \gamma] = \alpha\, \mathbb{E}[v_1] + \beta\, \mathbb{E}[v_2] + \gamma$

v.c.

$\in \mathbb{R}$

$$= \mathbb{E}[\alpha v_1] + \mathbb{E}[\beta v_2] + \mathbb{E}[\gamma]$$

- VARIANZA di una v.c. continua è:

$$Var[v] = \int_{-\infty}^{+\infty} \boxed{(x - \mathbb{E}[v])^2}\, f(x)\, dx$$

$$= \mathbb{E}\left[(v - \mathbb{E}[v])^2\right]$$

Osservazione

- $Var[v] \geq 0$. Se $Var[v] = 0 \Rightarrow v$ è deterministica
- Deviazione standard: $\sigma[v] = \sqrt{Var[v]}$
- $Var[v] = \mathbb{E}\left[(v - \mathbb{E}[v])^2\right] = \mathbb{E}\left[v^2 - 2\mathbb{E}[v]v + \mathbb{E}[v]^2\right]$

$$= \mathbb{E}[v^2] - 2 \cdot \mathbb{E}\left[\boxed{\mathbb{E}[v]} \cdot v\right] + \mathbb{E}\left[\boxed{\mathbb{E}[v]^2}\right]$$

$$= \mathbb{E}\left[v^2\right] - 2\cdot\mathbb{E}\left[\mathbb{E}[v]\cdot v\right] + \mathbb{E}\left[\mathbb{E}[v]^2\right]$$

$$= \mathbb{E}\left[v^2\right] - 2\,\mathbb{E}[v]\cdot\mathbb{E}[v] + \mathbb{E}[v]^2 = \boxed{\mathbb{E}\left[v^2\right] - \mathbb{E}[v]^2}$$

- $Var\left[\alpha v_1 + \beta\right] = \alpha^2 \cdot Var[v_1]$

- Date due v.c. v_1 e v_2 si definisce il COEFFICENTE DI CORRELAZIONE.

$$\rho = \frac{\mathbb{E}\left[\left(v_1 - \mathbb{E}[v_1]\right)\cdot\left(v_2 - \mathbb{E}[v_2]\right)\right]}{\sigma[v_1]\cdot\sigma[v_2]}$$

ρ indica il grado di dipendenza lineare tra due v.c. Se $\rho = 0$, le due variabili si dicono SCORRELATE

- Date due v.c. v_1 e v_2 si definisce la COVARIANZA come

$$Cov(v_1, v_2) = \mathbb{E}\left[\left(v_1 - \mathbb{E}[v_1]\right)\cdot\left(v_2 - \mathbb{E}[v_2]\right)\right]$$

$$\rho = \frac{Cov(v_1, v_2)}{\sigma[v_1]\cdot\sigma[v_2]}$$

Le precedenti definizioni possono essere estese al caso MULTIVARIABILE

$\Downarrow$

Vettore di v.c. $\bar{v} = \left[v_1 \quad v_2 \quad \text{—} \quad v_d\right]^T \in \mathbb{R}^{d\times 1}$

- distribuzione di probabilità

$$F(z_1, z_2, \text{—}, z_d) = P(v_1 \le z_1, v_2 \le z_2, \text{—}, v_d \le z_d)$$

$$= \int_{-\infty}^{z_1}\int_{-\infty}^{z_2} \text{—} \int_{-\infty}^{z_d} \underbrace{f(x_1, x_2, \text{—}, x_d)}_{\text{pdf CONGIUNTA}}\, dx_1\, dx_2 \text{—} dx_d$$

- Valore atteso:

$$\mathbb{E}\left[\bar{v}\right] = \left[\mathbb{E}[v_1],\ \mathbb{E}[v_2],\ \text{—},\ \mathbb{E}[v_d]\right]^T \in \mathbb{R}^{d\times 1}$$

$$\mathbb{E}[\bar{v}] = \big[\, \mathbb{E}[v_1],\ \mathbb{E}[v_2],\ —,\ \mathbb{E}[v_d] \,\big] \in \mathbb{R}^{d \times 1}$$

- Matrice di varianze - covarianze

$\bar{x} = [x_1\ x_2\ —\ x_d]^T \in \mathbb{R}^{d \times 1}$

$$\mathrm{Var}[\bar{v}] = \int_{\mathbb{R}^d} \underbrace{(\bar{x} - \mathbb{E}[\bar{v}])}_{d \times 1} \underbrace{(\bar{x} - \mathbb{E}[\bar{v}])^T}_{1 \times d} \cdot f(\bar{x})\, d\bar{x}$$

$$\Sigma = \underbrace{\begin{bmatrix} \mathrm{Var}[v_1] & \mathrm{Cov}(v_1, v_2) & & \\ \mathrm{Cov}(v_2, v_1) & \mathrm{Var}[v_2] & & \\ & & \ddots & \\ & & & \mathrm{Var}[v_d] \end{bmatrix}}_{d \times d}$$

- SIMMETRICA
- SEMIDEFINITA POSITIVA

$$\underset{1 \times d}{z^T} \underset{d \times d}{\Sigma} \underset{d \times 1}{z} \geq 0 \qquad \forall z \in \mathbb{R}^{d \times 1}$$

- Due variabili casuali v_1 e v_2 con pdf congiunta $f(x_1, x_2)$ si dicono INDIPENDENTI se e solo se:

$$f(x_1, x_2) = f(x_1) \cdot f(x_2)$$

Se v_1 e v_2 sono INDIPENDENTI $\Rightarrow$ sono SCORRELATE

STIMA 2

STIMA PARAMETRICA: stimare un vettore di parametri $\theta \in \mathbb{R}^{d \times 1}$

DATI: $D = \{ y(1), y(2), —, y(N) \}$ N. numero di dati osservati

Esempio LANCIO MONETA $D = \{ 1, 1, 0, 1, 0, 0 — 1 \}$

$\Downarrow$

θ°: PROBABILITÀ CHE ESCA TESTA

DATI (IN GENERALE): $D(s, \theta^\circ)$

DATI OSSERVATI: $D(\bar{s}, \theta^\circ)$

$$\theta = [\mu, \sigma^2] \qquad \theta^\circ = [\underset{\downarrow\ 5}{\mu^\circ}, \underset{\downarrow\ 3,2}{\sigma^{2,\circ}}]$$

Uno STIMATORE è una funzione $T(D(s, \theta^\circ))$. La STIMA è il risultato di uno stimatore su una specifica realizzazione dei dati

$$\hat{\theta} = T(D(\bar{s}, \theta^\circ))$$

s_1 = primi 10 studenti "estratti" $\quad T(D(s_1, \theta^\circ)) = \hat{\theta}_1$

s_1 = prima lo studenti "estratti" $T(D(s_1, \vartheta^\circ)) = \hat{\vartheta}_1$

s_2 = secondi " " " $T(D(s_2, \vartheta^\circ)) = \hat{\vartheta}_2 \neq \hat{\vartheta}_1$

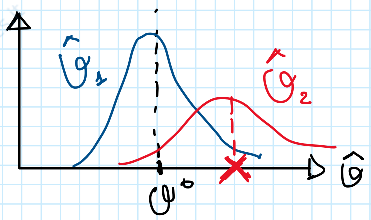
Lo STIMATORE, poiché il suo valore dipende dall'esito s, allora può essere interpretato come una v.c.

$\Downarrow$

Ha senso parlare di VALORE ATTESO e VARIANZA dello stimatore

* PROPRIETÀ DI UNO STIMATORE * $\boxed{2.1}$

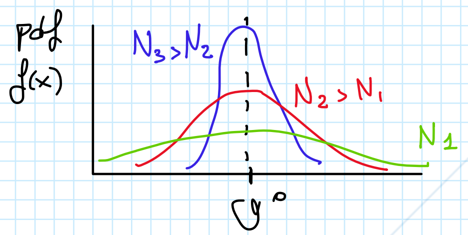
- CORRETTO se e solo se $\mathbb{E}[\hat{\vartheta}] = \vartheta^\circ$



- ASINTOTICAMENTE CORRETTO se e solo se $\lim_{N \to +\infty} \mathbb{E}[\hat{\vartheta}] = \vartheta^\circ$

- CONSISTENTE se e solo se: $\forall \varepsilon > 0$

$$\lim_{N \to +\infty} P(|\overbrace{T(\cdot)}^{\hat{\vartheta}} - \vartheta^\circ| > \varepsilon) = 0$$



Se ho due (o più) stimatori corretti, quale scelgo? Quello che varianza minima

- OTTIMO: uno stimatore corretto si dice ottimo s.s.e la sua varianza è la più piccola per una serie di dati $D = \{y(1), \_\_, y(N)\}$ di lunghezza N

Esempio: STIMATORE MEDIA $\vartheta = \mu$ $\vartheta = [\mu, \sigma^2]$

Siano $D = \{y(1), \_\_, y(N)\}$ variabili casuali con media μ e varianza σ^2.
Lo stimatore MEDIA CAMPIONARIA $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} y(i)$ è corretto

$\mathbb{E}[\hat{\mu}] = \mu$?

$$\mathbb{E}[\hat{\mu}] = \mu \ ?$$

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} y(i)\right] = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N} y(i)\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[y(i)]$$

$$= \frac{1}{\cancel{N}}\cdot \cancel{N}\cdot \mu = \mu$$

$$= \frac{1}{N}\Big(\mathbb{E}[y(1)] + \mathbb{E}[y(2)] + \dots + \mathbb{E}[y(N)]\Big)$$

$$= \frac{1}{N}(\mu + \mu + \dots + \mu)$$

$$= \mu$$

Si dimostra che $\hat{\mu}$ è CONSISTENTE

$$Var[\hat{\mu}] = \frac{Var[y]}{N} = \frac{\sigma^2}{N}$$

LIMITE DI CRAMER - RAO 2.2

stabilisce un limite inferiore alla varianza di qualsiasi stimatore

Nel caso di STIMATORI CORRETTI abbiamo che $\underset{1\times 1}{Var[\hat{\vartheta}]} \geq m^{-1}$ — ϑ SCALARE

m: quantità di informazione di Fisher

se $\hat{\vartheta}$ è un vettore: $\underset{d\times d}{Var[\hat{\vartheta}]} - \underset{d\times d}{M^{-1}} \geq 0$

- EFFICIENTE: sse $Var[\hat{\vartheta}] = m^{-1}$
- ASINTOTICAMENTE EFFICIENTE: sse $\lim_{N\to+\infty} Var[\hat{\vartheta}] = m^{-1}$

STIMA A MINIMI QUADRATI (LEAST SQUARES) 3

Partendo dai dati $D = \{y(1), \dots, y(N)\}$, li abbiamo descritti in termini della loro media e varianza

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y(i) \qquad \hat{S}^2_{N-1} = \frac{\sum_{i=1}^{N}(y(i)-\hat{\mu})^2}{N-1}$$

MEDIA CAMPIONARIA — VARIANZA CAMPIONARIA

Supponiamo di voler descrivere i dati D tramite una relazione lineare

$x_0(i)$

$i = 1 \dots N$

$$y(i) = \vartheta_0 \cdot 1 + \vartheta_1 x_1(i) + \vartheta_2 x_2(i) + \dots + \vartheta_{d-1}x_{d-1}(i) + e(i)$$

NOTE — REGRESSORE 1 — NOTI — ERRORE

$$\vartheta = [\vartheta_0 \ \vartheta_1 \ \dots \ \vartheta_{d-1}]^T \in \mathbb{R}^{d\times 1}$$

NOTE

NOTI

REGRESSORE 1
COVARIATA 1
FEATURE 1

$$\underset{d\times 1}{\theta} = \left[\theta_0 \ \ \theta_1 \ \ - \ \ \theta_{d-1}\right]^T \in \mathbb{R}^{d\times 1}$$

Esempio

y : PESO [kg]

x_1 : ALTEZZA [m]

x_2 : ETÀ [anni]

x_3 : SESSO [M/F]

16 PERSONE

REGRESSORE 2

ALTEZZA	ETÀ	SESSO	PESO
170	32	M	60
185	18	F	80

1 SOGGETTO

$N = 16$

REGRESSORE 1

$d-1$

Definiamo i vettori:

$$\underset{d\times 1}{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ | \\ \theta_{d-1} \end{bmatrix} \qquad \underset{d\times 1}{\varphi(i)} = \begin{bmatrix} 1 \\ x_1(i) \\ x_2(i) \\ | \\ x_{d-1}(i) \end{bmatrix} \Rightarrow$$

$$\boxed{\begin{gathered} i = 1 - N \\ \underset{1\times 1}{y(i)} = \underset{1\times d}{\varphi(i)^T} \cdot \underset{d\times 1}{\theta} + e(i) \end{gathered}}$$

Il metodo dei MINIMI QUADRATI minimizza:

$$J(\underset{d\times 1}{\theta}) = \frac{1}{N} \cdot \sum_{i=1}^{N} \Big(y(i) - \underbrace{\varphi(i)^T \cdot \theta}_{\hat{y}(i)} \Big)^2 = \frac{1}{N} \sum_{i=1}^{N} e(i)^2$$

$$\Rightarrow \underset{1\times 1}{\alpha} \cdot \underset{1\times 1}{\theta} \Rightarrow \frac{d(\alpha\theta)}{d\theta} = \alpha$$

$$\underset{d\times 1}{\hat{\theta}} = \arg\min_{\theta} J(\theta)$$

$$\underset{d\times 1}{\nabla_\theta J(\theta)} = \underset{d\times 1}{0} \Rightarrow \cancel{\frac{-1\cdot 2}{N}} \cdot \sum_{i=1}^{N} \underset{d\times 1}{\varphi(i)} \Big(\underset{1\times 1}{y(i)} - \underset{1\times d}{\varphi(i)^T} \cdot \underset{d\times 1}{\theta} \Big) = 0$$

$$\sum_{i=1}^{N} \varphi(i) \cdot y(i) - \sum_{i=1}^{N} \varphi(i)\varphi(i)^T \theta = 0$$

$$\underbrace{\left[\sum_{i=1}^{N} \underset{d\times 1}{\varphi(i)} \underset{1\times d}{\varphi(i)^T} \right]}_{d\times d} \cdot \underset{d\times 1}{\theta} = \sum_{i=1}^{N} \underset{d\times 1}{\varphi(i)} \cdot y(i)$$

FORMA CHIUSA

$$\hat{\theta} = \left[\sum_{i=1}^{N} \varphi(i) \cdot \varphi(i)^T \right]^{-1} \cdot \left[\sum_{i=1}^{N} \varphi(i)\, y(i) \right]$$

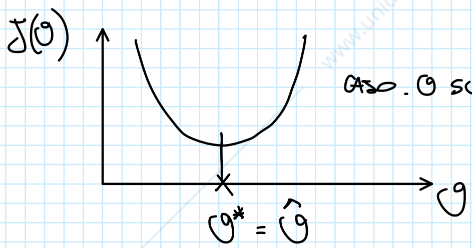
$$\underset{d\times 1}{\hat{\theta}} = \Big[\underbrace{\sum_{i=1}^{N} \varphi(i)\cdot\varphi(i)^T}_{d\times d}\Big]^{-1} \cdot \Big[\underbrace{\sum_{i=1}^{N} \varphi(i)\, y(i)}_{d\times 1}\Big]$$

Osservazione

- Se $\det\Big[\sum_{i=1}^{N} \varphi(i)\varphi(i)^T\Big] \neq 0 \Rightarrow$ la soluzione è UNICA!
- Se $\quad$ // $\quad = 0 \Rightarrow \exists$ infinite soluzioni
- Dato che il modello è lineare (nei parametri) e la funzione di costo $J(\theta)$ è quadratica (nei parametri)

$\Downarrow$

Allora $J(\theta)$ ha UN UNICO MINIMO (globale)



CASO θ SCALARE