COMPITO DI
IDENTIFICAZIONE DEI MODELLI ED ANALISI DEI DATI
23 luglio 2012

Domanda 1. Si assuma che un sistema fisico di interesse con ingresso $u(t)$ (misurato) ed uscita $y(t)$ si possa modellare tramite una funzione di trasferimento (assegnata) $H(z)$, cioè tale che $Y(z) = H(z)U(z)$. Si vuole costruire un sistema di controllo tramite feedback (lineare) dallo stato. Tuttavia lo stato NON è accessibile ma sono disponibili misure rumorose $z(t) = y(t)+d(t)$ dove il disturbo $d(t)$ è un processo stazionario, a media nulla e scorrelato da $u(t)$, che si può modellare come l'uscita di un sistema lineare BIBO stabile con funzione di trasferimento $W(z)$ (nota) ed ingresso un processo $n(t)$ di rumore bianco.

Si descriva nel maggior dettaglio possibile come utilizzare l'algoritmo del Filtro di Kalman per ottenere una stima dello stato $x(t)$ (lo stato di una opportuna realizzazione della funzione di trasferimento $H(z)$.)

Supponendo che la funzione di trasferimento $W(z)$ sia nota a meno di una costante moltiplicativa c, si illustri come sia possibile "tarare" la costante c in base al funzionamento del filtro di Kalman.

Domanda 2. Si assuma che il processo $y(t)$ sia generato dal modello dinamico

$$y(t) = a_1^0 y(t-1) + b_1^0 u(t-1) + e_0(t) \quad t \in \mathbb{Z}$$

dove $e_0(t)$ è un rumore bianco, a media nulla e varianza σ^2, indipendente dal processo u. Il processo di "ingresso" $u(t)$ si assume bianco, a media nulla e varianza unitaria.

Per fare "identificazione", si utilizzi il criterio di minimizzazione dell'errore di predizione (PEM) ed il modello

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t)$$

con $e(t)$ bianco.

1. Si calcoli la varianza del processo $y(t)$; si enunci una condizione sufficiente sui coefficienti a_1^0, b_1^0 affinchè il processo $y(t)$ sia stazionario e a varianza finita.

2. Assumendo siano disponibili dati $y(t), u(t)$, $t \in [1, N]$, si imposti il problema di identificazione; si trovi lo stimatore $\hat{\underline{\theta}}_N$ del parametro vettoriale $\underline{\theta} := [b_0 \; b_1]$, scrivendo chiaramente quale funzione dei dati si deve minimizzare.

3. Si dica se lo stimatore $\hat{\theta}_N$ PEM converge e se si, dove.

Domanda 3. Si assuma di avere un'urna contenti due tipi di dadi (diciamo A e B) indistinguibili alla vista. Ci sono 5 dadi di tipo A e 10 dadi di tipo B. I dadi di tipo A sono equi mentre i dadi di tipo B sono truccati e forniscono il risultato 6 con probabilità $1/3$ ed una qualunque altra faccia con probabilità $\frac{2}{15}$. Si assuma di estrarre un dado a caso e di lanciarlo 6 volte, ottenendo la sequenza $1, 3, 6, 6, 2, 6$.

Si calcoli:

1. Indicando con X l'esito di un lancio del dato estratto si calcoli la probabilità $P[X = i]$, $i = 1, .., 6$.

2. Si calcoli la probbailità che il dado estratto sia il dado A sapendo che l'esito di 6 lanci è quello riportato sopra

Domanda 4. Si voglia stimare la funzione $f(x) : \mathbb{R} \to \mathbb{R}$ a partire da misure rumorose

$$y_i = f(x_i) + e_i \quad i = 1, .., N$$

Si assuma che e_i siano rumori Gaussiani indipendenti a media zero e varianza σ^2 *nota*. La funzione incognita $f(x)$ viene descritta nel seguente modo: per ogni scelta di punti x_j, $j = 1, .., M$ le variabili $z_i := f(x_i)$ sono tali che

$$z := [z_1, z_2, .., z_M]$$

è un vettore Gaussiano, a media zero e matrice covarianza C, tale che il suo elemento in posizione (i, j) è $C_{x_i,x_j} = cov(f(x_i), f(x_j)) = \mathbb{E}f(x_i)f(x_j)$.

Si assuma inoltre che il rumore e_i è indipendente da $f(x)$ per ogni scelta di x ed i.

1. Dato un nuovo punto $x \in \mathbb{R}$ si ricavi lo stimatore a minima varianza di $f(x)$ e la sua varianza.
2. Cosa si può dire se $x = x_i$ è uno dei punti $i \in 1, ..N$ nei quali si aveva fatto delle misure?
3. Assumendo che[1] $C(x_i, x_j) = e^{\frac{-(x_i - x_j)^2}{\gamma}}$; pensate sia possibile stimare γ dai dati e, se si, come?

[1] C'era un errore di stampa nel testo del compito, che comunque non cambiava la metodologia di soluzione.

SOLUZIONI

Domanda 1.[8 Punti] Si tratta di impostare il problema di stima dello stato utilizzando il filtro di Kalman. A questo scopo è sufficiente procurarsi una realizzazione minima (A_H, B_H, C_H, D_H) della funzione di trasferimento $H(z) = C_H(zI - A_H)^{-1}B_H + D_H$ e una realizzazione minima (A_W, B_W, C_W, D_W) della funzione di trasferimento $W(z) = C_W(zI - A_W)^{-1}B_W + D_W$. Dalle ipotesi del problema si ricava immediatamente che

$$\begin{array}{rcl} \begin{bmatrix} x(t+1) \\ s(t+1) \end{bmatrix} & = & \begin{bmatrix} A_H & 0 \\ 0 & A_W \end{bmatrix} \begin{bmatrix} x(t) \\ s(t) \end{bmatrix} + \begin{bmatrix} B_H \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ B_W \end{bmatrix} n(t) \\ z(t) & = & [C_H \;\; C_W] \begin{bmatrix} x(t) \\ s(t) \end{bmatrix} + D_H u(t) + D_W n(t) \end{array} \tag{1}$$

dove si assume, senza perdita di generatià, che il rumore bianco $n(t)$ di ingresso abbia varianza unitaria.

Le equazioni in (1) forniscono una realizzazione di stato per il processo (stazionario) di misura $z(t)$ con stato $\xi(t) := [x^\top(t) \; s^\top(t)]^\top$ dove $x(t)$ rappresenta lo stato di una realizzazione minima di $H(z)$ mentre $s(t)$ è lo stato di una realizzazione minima del modello del rumore di misure. Si può applicare quindi l'algoritmo del filtro di Kalman (nel compito è opportuno riportare le equazioni, si vedano gli appunti) con

$$A := \begin{bmatrix} A_H & 0 \\ 0 & A_W \end{bmatrix} \quad B := \begin{bmatrix} B_H \\ 0 \end{bmatrix}$$

$$C := [C_H \;\; C_W] \quad D = D_W$$

e

$$Q = BB^\top \quad R = DD^\top \quad S = BD^\top$$

per ottenere una stima $\hat{\xi}(t|t)$ dello stato $\xi(t)$ da cui poi $\hat{x}(t|t) = [I \; 0]\hat{\xi}(t|t)$.

Dire che la funzione di trasferimento è nota solamente e meno di una costante moltiplicativa è come dire che le matrici B_W e D_W che modellano il rumore sono note a meno di una costante moltiplicativa comune o, equivalentemente, che il rumore di ingresso $n(t)$ ha varianza incognita $Var\{n(t)\} = \gamma$

Si ricordi che, con la notazione del filtro di Kalman utilizzata a lezione abbiamo $Q = \gamma BB^\top$, $R = \gamma DD^\top$ e $S = \gamma BD^\top$. Di conseguenza si può tarare l'algoritmo del filtro di Kalman variando il valore γ fin a trovare il valore per il quale l'errore di predizione $z(t) - \hat{z}(t|t-1)$ è bianco; quest'ultima condizione si può ad esempio verificare utilizzando il test del periodogramma cumulato.

Domanda 2.[8 Punti]

1. Il processo di uscita è stazionario se e solo se il modello è asintoticamente stabile, cioè se $|a_1^0| < 1$ (le risposte impulsive da u a y e da e_0 a y devono essere sommabili al quadrato, e visto che sono le risposte impulsive di un sistema descritto da un'equazione alle differenze, il sistema di partenza deve essere asintoticamente stabile (si noti che non ci sono cancellazioni in generale) ; si noti che dalle ipotesi discende che $\mathbb{E}y_t = 0$. La varianza del processo di uscita y_t si ottiene semplicemente dalla relazione:

   $$\mathbb{E}y^2(t) = \mathbb{E}\left(a_1^0 y(t-1) + b_1^0 u(t-1) + e_0(t)\right)^2$$

   dalle ipotesi dell'esercizio u è completamente scorrelato da e_0 e, in aggiunta, poichè $y(t-1)$ è un funzionale lineare e stabile del passato di u e e_0, cioè di $u(s)$ per $s < t-1$ e di e $e_0(s)$ per $s \leq t-1$, allora $y(t-1)$ è scorrelato da $u(t-1)$ ed $e_0(t)$ (perchè sia u che e_0 sono bianchi) e quindi i tre termini del membro di destra dell'ultima equazione sono scorrelati. Di conseguenza

$$\mathbb{E}y^2(t) = (a_1^0)^2 \mathbb{E}y^2(t-1) + (b_1^0)^2 \mathbb{E}u^2(t-1) + \mathbb{E}e_0^2(t)$$

Poichè il processo è stazionario si ha $\sigma_y^2 = \mathbb{E}y^2(t) = \mathbb{E}y^2(t-1)$ e quindi:

$$\sigma_y^2 = (a_1^0)^2 \sigma_y^2 + (b_1^0)^2 \mathbb{E}u^2(t-1) + \mathbb{E}e_0^2(t) = (a_1^0)^2 \sigma_y^2 + (b_1^0)^2 \ \ + \sigma^2$$

da cui:

$$\sigma_y^2 = \frac{(b_1^0)^2 \ \ + \sigma^2}{1 - (a_1^0)^2}.$$

2. Il modello scelto per fare identificazione è lineare nei parametri e si può scrivere nella forma

$$y(t) = \phi^\top(t)\theta + e(t)$$

dove $\phi^\top(t) := [u(t)\, u(t-1)]$; e $\theta^\top = [b_0, b_1]$. Utilizzando il metodo di minimizzazione dell'errore di predizione

$$\begin{aligned} \hat{\theta}_N &:= \arg\min_\theta \tfrac{1}{N} \textstyle\sum_{t=1}^N (y(t) - \hat{y}_\theta(t|t-1))^2 \\ &= \arg\min_\theta \tfrac{1}{N} \textstyle\sum_{t=1}^N (y(t) - \phi^\top(t)\theta)^2 \end{aligned}$$

la cui soluzione è data da:

$$\hat{\theta}_N := \left[\sum_{t=1}^N \phi(t)\phi^\top(t)\right]^{-1} \sum_{t=1}^N \phi(t)y(t)$$

3. Lo stimatore PEM converge, in ipotesi di ergodicità, all'insieme dei punti di minimo del funzionale

$$J(\theta) := \mathbb{E}(y(t) - \phi^\top(t)\theta)^2 = \mathbb{E}\left(a_1^0 y(t-1) + (b_1^0 - b_1)u(t-1) - b_0 u(t) + e_0(t)\right)^2$$

Tutte le variabili aleatorie al secondo membro di questa equazione sono variabili scorrelate (per lo stesso motivo descritto nella soluzione del punto 1) e quindi

$$J(\theta) = b_0^2 + (a_1^0)^2 \sigma_y^2 + (b_1^0 - b_1)^2 + \sigma^2$$

che, come funzione di b_0 e b_1 ha chiaramente un unico punto di minimo per $(b_0, b_1) = (0, b_1^0)$.

Domanda 3.[8 Punti] Indichiamo con E_A l'evento *il dado estratto dall'urna è di tipo A* e con E_B l'evento *il dato estratto dall'urna è di tipo B*. Chiaramente la probabilità $P[E_A] = \frac{5}{15} = \frac{1}{3}$ e $P[E_B] = \frac{10}{15} = \frac{2}{3}$.

1. Per calcolare la probabilità che l'esito X del dado estratto dall'urna sia uguale alla faccia i si può usare il Teorema della Probabilità totale:

$$P[X = i] = P[X = i|E_A]P[E_A] + P[X = i|E_B]P[E_B]$$

Di conseguenza, usando i dati del problema si ottiene:

$$P[X = i] = \begin{cases} \bar{p}_6 = \frac{1}{6}\frac{1}{3} + \frac{2}{15}\frac{2}{3} = \frac{13}{90} & i = 1, 2, 3, 4, 5 \\ p_6 = \frac{1}{6}\frac{1}{3} + \frac{1}{3}\frac{2}{3} = \frac{25}{90} & i = 6 \end{cases}$$

Si noti che $5\bar{p}_6 + p - 6 = 1$.

2. Si indichi con $Y := [X_1, .., X_6]$ il vettore che modella i sei lanci del dado estratto e sia $y = [x_1, x_2, .., x_6] = [1\ 3\ 6\ 6\ 2\ 6]$ il vettore dei sei esiti. si vuole calcolare $P[E_A|Y = y]$; utilizzando la regola di Bayes ed il Teorema della Probabilità Totale

$$P[E_A|Y = y] = \frac{P[Y = y|E_A]P[E_A]}{P[Y = y]} = \frac{P[Y = y|E_A]P[E_A]}{P[Y = y|E_A]P[E_A] + P[Y = y|E_B]P[E_B]}$$

Si noti ora che, condizionatamente agli eventi E_A o E_B, le variabili X_1, X_2, .., X_6 sono indipendenti (perchè) sono lanci indipendenti dello stesso dado ed il condizionamento identifica la sua legge di probabilità (equo/truccato). Quindi:

$$P[Y = y|E_A] = \prod_{i=1}^{6} P[X_i = x_i|E_A] = \left(\frac{1}{6}\right)^6$$

$$P[Y = y|E_B] = \prod_{i=1}^{6} P[X_i = x_i|E_B] = \left(\frac{1}{3}\right)^3 \left(\frac{2}{15}\right)^3$$

Dove il risultato discende dal fatto che, se il dado è equo tutti gli esiti sono equiprobabili ed hanno probabilità $\frac{1}{6}$, mentre nel caso del dado truccato la faccia 6 (che esce tre volte) ha probailità $\frac{1}{3}$ mentre le altre facce sono equiprobabili e hanno probabilità $\frac{2}{15}$. Di conseguenza la probabilità cercata vale:

$$P[E_A|Y = y] = \frac{\left(\frac{1}{6}\right)^6 \frac{1}{3}}{\left(\frac{1}{6}\right)^6 \frac{1}{3} + \left(\frac{1}{3}\right)^3 \left(\frac{2}{15}\right)^3 \frac{2}{3}} = \frac{1}{1 + 2^6 * 2\left(\frac{2}{5}\right)^3} \simeq 0.1088$$

Domanda 4.[8 Punti]

Le ipotesi dell'esercizio comunque si scelgano dei punti $x_1, ..., x_M$ le variabili $f(x_1), ..., f_(x_M)$ sono congiuntamente Gaussiane e indipendenti da e. Poiche $y_1, .., y_N$ sono ottenute come somma di variabili Gaussiane e indipendenti sono ancora Gaussiane. Di conseguenza, per ogni scelta di x, le variabili $f(x), y_1, .., y_N$ sono congiuntamente Gaussiane. Si indichi con $C(\alpha, \beta) := \mathbb{E} f(\alpha)f(\beta)$. Con questa premessa rispondiamo alle domande dell'esercizio:

1. Poichè $f(x), y_1, .., y_N$ sono congiuntamente Gaussiane lo stimatore a minima varianza di $f(x)$ date le misure $y_1, .., y_N$ (cioè la media condizionata di $f(x)$ dati $y := (y_1, .., y_N)^\top$) si calcola in forma chiusa con la formula (tutte le variabili in gioco hanno media nulla per ipotesi)

$$\hat{f}(x) = \mathbb{E}[f(x)|y_1, ..y_N] = cov(f(x), Y)Var\{y\}^{-1}y = [cov(f(x), y_1), ..., cov(f(x), y_N]\, Var\{y\}^{-1}y$$

Per calcolare $cov(f(x), y_i)$ basta procedere come segue:

$$cov(f(x), y_i) = cov(f(x), f(x_i){+}e_i) = cov(f(x), f(x_i)){+}cov(f(x), e_i) = cov(f(x), f(x_i)) = C(x, x_i)$$

mente gli elementi della matrice varianza $Var\{y\}$ si ottengono dalla relazione

$$Var\{y\}_{ij} = cov(y_i, y_j) = cov(f(x_i){+}e_i, f(x_j){+}e_j) = cov(f(x_i), f(x_j)){+}\sigma^2\delta_{ij} = C(x_i, x_j){+}\sigma^2\delta_{ij}$$

dove di è usata come sopra l'indipendenza tra e_i e $f(x)$ ($\forall x$) e δ_{ij} è il simbolo di Kronecker, che vale 1 se e solo se $i = j$ e 0 altrimenti. La varianza dello stimatore $\hat{f}(x)$ si ottiene immediatamente dalla teoria nota nel caso di variabili congiuntamente Gaussiane fornendo:

$$Var\{f(x)|Y\} = Var f(x) - cov(f(x), Y)Var\{y\}^{-1}cov(f(x), Y)^\top$$

2. Si osservi che se x è uno dei punti in cui si sono fatte misure (assumiamo $x = x_1$ senza perdita di generalità), si sta cercando di stimare il valore di $f(x_1)$ utilizzando tutte le misure. Un altro stimatore possibile di $f(x_1)$ è quello basato sulla sola misura y_1. in questo caso, ovviamente
$$\mathbb{E}[f(x_1)|y_1] = \frac{C(x_1,x_1)}{C(x_1,x_1)+\sigma^2} y_1$$
la cui varianza è
$$Var\{f(x_1)|y_1\} = C(x_1,x_1) - \frac{C(x_1,x_1)^2}{C(x_1,x_1)+\sigma^2} = \frac{C(x_1,x_1)}{C(x_1,x_1)+\sigma^2}\sigma^2 < \sigma^2$$
Si osservi che vale la seguente catena di diseguaglianze:
$$Var\{f(x)|Y\} \leq Var\{f(x_1)|y_1\} = \frac{C(x_1,x_1)}{C(x_1,x_1)+\sigma^2}\sigma^2 < \sigma^2$$
dove la prima diseguaglianza discende dal fatto che lo stimatore ottimo che usa *tutte* le misure deve comportarsi meglio dello stimatore che usa solo y_1. L'ultima diseguaglianza, che e' una immediata conseguenza del fatto che $\frac{C(x_1,x_1)}{C(x_1,x_1)+\sigma^2} < 1$, mostra che lo stimatore "Bayesiano" ha una varianza pi piccola dell'errore di misura. Cioè combinando l'informazione a priori con le misure si riesce ad ottenere una stima finale che è migliore di quella che si sarebbe ottenuta utilizzando semplicemente la misura y_1 come stimatore di $f(x_1)$ (la cui varianza sarebbe stata pari alla varianza del rumore di cui era affetta la misura y_1.

3. Si indichi con $C_\gamma(\alpha,\beta) = cov\{f(\alpha), f(\beta)\} = e^{-\frac{(\alpha-\beta)^2}{\gamma}}$. Per γ fissato la misure sono campioni di variabili aleatorie congiuntamente Gaussiane con media zero e matrice covarianza che dipende dal valore di γ, i.e. $\Sigma(\gamma) := \mathbb{E}yy^\top$ il cui elemento in posizione i,j vale:
$$\Sigma_{ij}(\gamma) = cov(y_i, y_j) = C_\gamma(x_i, x_j) + \sigma\delta_{ij}$$
e si Di conseguenza la densità di probabilità delle misure $y = [y_1, .., y_N]^\top$ ha la forma
$$f_\gamma(y) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma(\gamma))}} e^{-\frac{1}{2}\ y^\top \Sigma^{-1}(\gamma) y}$$
Non appena sono disponbili le misure è possibile simare γ usando l criterio della massima verosimiglianza
$$\hat{\gamma}_{ML} := \arg\ \max_{\gamma>0} \mathbb{F}_\gamma(y)$$
La funzione $f_\gamma(y)$ si chiama "likelihood marginale" e viene comunemente usata in questo modo per stimare il valore di γ (la covarianza assegnata da luogo a uno stimatore che si chiama nella letteratura "rete di regolarizzazione")