COMPITO DI
IDENTIFICAZIONE DEI MODELLI ED ANALISI DEI DATI
3 settembre 2012

Domanda 1. Sia $y(t) \in \mathbb{R}$, $t \in \mathbb{Z}$, un processo descritto da un modello $ARMA(n, n)$ del tipo

$$\sum_{k=0}^{n} a_k y(t-k) = \sum_{k=0}^{n} c_k e(t-k) \quad t \in \mathbb{Z}$$

dove $e(t)$ è un rumore stazionario, Gaussiano, bianco e a media nulla. L'equazione alle differenze si deve intendere risolta "in avanti", cioè considerando il sistema causale.

1. Si dica sotto quali condizioni (sui coefficienti a_k e c_k) il processo $y(t)$ è stazionario.
2. Si calcoli la media di $y(t)$, il suo spettro e si indichi un procedimento per calcolare la funzione di covarianza $cov(y(t+\tau)y(t))$.
3. Si descriva un procedimento per calcolare il predittore lineare a minima varianza $\hat{y}(t|t-1)$.

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

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + e_0(t) \quad t \in \mathbb{Z}$$

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

$$u(t) = v(t) + v(t-1)$$

dove $v(t)$ è bianco, a media nulla e varianza unitaria. Per fare "identificazione", si utilizzi il criterio di minimizzazione dell'errore di predizione (PEM) ed il modello

$$y(t) = \theta u(t-1) + e(t)$$

dove si assume $e(t)$ bianco.

1. Si calcoli la varianza del processo $y(t)$;
2. Assumendo siano disponibili dati $y(t), u(t)$, $t \in [1, N]$, si imposti il problema di identificazione; si trovi lo stimatore $\hat{\theta}_N$ del parametro θ, 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 voler progettare un sistema di visione per trovare in un'immagine (I d'ora in poi), in maniera automatica, quale di due oggetti è presente[1]. Ad esempio i due oggetti potrebbero essere un pallone da calcio (oggetto A) e un pallone da basket (oggetto B). Il sistema elabora un'immagine I ed estrae k "features" [2] (chiamiamole $y_i = g_i(I)$, $y_i \in \mathbb{R}$, $i = 1, .., k$). Un classificatore è una funzione

$$c(y_1, .., y_k) : \underbrace{(y_1, ..., y_k)}_{\in \mathbb{R}^k} \to \underbrace{\hat{C}}_{\in \{A,B\}}$$

[1] Per semplicità assumiamo che (i) ci siano solo due oggetti "di interesse" e che, sicuramente, (ii) uno e solo uno dei due sia presente in ogni immagine.

[2] Si chiamano "features" delle funzioni $g(I)$, arbitrarie in principio, dell'immagine, ad esempio la media, la media dei valori in un piccola finestra dell'immagine intorno ad un pixel specifico etc...

che restituisce, per ogni $k - upla$ di features $(y_1, .., y_k) = (g_1(I), ...g_k(I))$, la stima dell'oggetto $\hat{C}$ (stima della "classe") che può essere uguale ad A (pallone da calcio) oppure a B (pallone da basket). ll classificatore può essere caratterizzato dalle seguenti probabilità:

$$P[c(y_1, .., y_k) = A|A] = p_A \quad P[c(y_1, .., y_k) = B|B] = p_B$$

che sono le probabilità di "decisione" corretta nei due casi in cui, rispettivamente, nell'immagine I ci sia un pallone da calcio o da basket. Ovviamente il classificatore può sbagliare, che succede con probabilità $P[c(y_1, .., y_k) = B|A]$ e $P[c(y_1, .., y_k) = A|B]$

1. Assumendo di estrarre a caso un'immagine da un database formato di 1000 immagini di cui un terzo contengano palloni da basket e le altre palloni da calcio, si calcoli la probabilià che la decisione finale sia sbagliata.

2. Assumendo che la "risposta" del classificatore su una particolare immagine sia $\hat{C} = c(y_1, .., y_k) = A$, si calcoli la probabilità che sia effettivamente presente un pallone da calcio.

3. Supponendo che sia nota la funzione $c(y_1, .., y_k)$, che siano fissate le funzioni $g_1, .., g_k$ usate per estrarre le features $(y_1, .., y_k)$ e che sia disponibile un database di immagini (diciamo I_n, $n = 1, ..., N$ con N grande) per cui è nota la classificazione (cioè si sa per ciascuna immagine se è presente un pallone da calcio o un pallone da basket). Sapreste indicare un procedimento per stimare le probabilità p_A e p_B?

Domanda 4. Si voglia stimare il percorso di un robot planare sul piano (x, y) in funzione del tempo t. La traiettoria si può descrivere tramite le coordinate $x(t), y(t)$, $t \in \mathbb{Z}$ (si assuma un modello a tempo discreto per semplicità, immaginando quindi di aver campionato con passo unitario la vera traiettoria a tempo continuo). Poiché il robot si muove in modo "continuo" il modello più semplice che si può immaginare è il seguente:

$$\begin{aligned} x(t+1) &= x(t) + r_x(t) \\ y(t+1) &= y(t) + r_y(t) \end{aligned}$$

dove $r_x(t)$ e $r_y(t)$ sono processi aleatori Gaussiani, indipendenti, a media nulla e funzione covarianza $cov(r_x(t), r_x(s)) = \sigma_r^2 \delta(t - s)$, $cov(r_y(t), r_y(s)) = \sigma_r^2 \delta(t - s)$; si assumono condizioni iniziali

$$x(0) = 0 \quad y(0) = 0.$$

Si assuma siano disponibili misure del tipo[3]

$$z_x(t_i) = x(t_i) + n_x(t_i) \quad z_y(t_i) = y(t_i) + n_y(t_i), \quad t_i \in [0, 1, 2, ..., T], i = 0, .., N$$

con $n_x(t)$ e $n_y(s)$ indipendenti $\forall t, s$, Gaussiani, a media nulle e funzione covarianza $cov(n_x(t), n_x(s)) = \sigma_n^2 \delta(t - s)$, $cov(n_y(t), n_y(s)) = \sigma_n^2 \delta(t - s)$.

1. Si ricavi l'espressione della funzione covarianza di $x(t)$ e $y(t)$, $t \in [0, T]$

2. Si ricavino le espressioni degli stimatori Bayesiani $\hat{\mathbb{E}}[x(t)|z_x(s), z_y(s), s \in [0, T]]$ e $\hat{\mathbb{E}}[y(t)|z_x(s), z_y(s), s \in [0, T]]$, $\forall t \in [0, T]$

3. Assumendo che siano disponibili misure per ogni istante di tempo intero nell'intervallo $[0, T]$ (e quindi che $N = T$) si illustri come sia possibile ricavare lo stimatore $\hat{\mathbb{E}}[x(t)|z_x(s), z_y(s), s \in [0, t]]$ e $\hat{\mathbb{E}}[y(t)|z_x(s), z_y(s), s \in [0, t]]$ in modo ricorsivo.

[3] Si noti che, in generale $N < T$, cioè non sono disponibili misure a tutti gli istanti ti tempo interi.

SOLUZIONI

Domanda 1.[8 Punti]

1. Il processo di uscita del modello ARMA causale è un processo stazionario se il modello ARMA è asintoticamente stabile, e quindi se il polinomio caratteristico
$$\sum_{k=0}^{n} a_k z^{n-k}$$
ha tutte le radici a modulo strettamente minore di uno; in realtà basterebbe la BIBO stabilità a patto che le condizioni iniziali siano nulle; quindi basta che la funzione di trasferimento
$$H(z) = \frac{\sum_{k=0}^{n} c_k z^{n-k}}{\sum_{k=0}^{n} a_k z^{n-k}}$$
non abbia poli in $|z| \geq 1$.

2. Per quanto rigurda la media di $y(t)$ basta osservare che
$$\mathbb{E}\left[\sum_{k=0}^{n} a_k y(t-k)\right] = \mathbb{E}\left[\sum_{k=0}^{n} c_k e(t-k)\right]$$
e usando la linearità del valore atteso:
$$\sum_{k=0}^{n} a_k \mathbb{E} y(t-k) = \sum_{k=0}^{n} c_k \mathbb{E} e(t-k) = 0$$
poichè $\mathbb{E}e(t) = 0$. Quindi la media $m(t) := \mathbb{E}y(t)$ soddisfa l'equazione omogenea associata
$$\sum_{k=0}^{n} a_k m(t-k) = 0$$
la cui soluzione ha la forma di una evoluzione libera, che tende a zero per l'asintotica stabilità (se il sistema fosse solo BIBO stabile si ricordi che si dovrebbero assumere condizioni iniziali nulle e quindi l'unica soluzione dell'omogenea associata sarebbe identicamente nulla). Quindi a regime $m(t) = 0$.

   Per calcolare la covarianza basta trovare lo spettro
$$\Phi_y(z) = H(z) H^\top(z^{-1}) \sigma_e^2$$
dove σ_e^2 è la varianza del rumore $e(t)$, e ricordare che la funzione di covarianza è l'antitrasformata zeta dello spettro
$$\Lambda(\tau) = \mathbb{E} y(t+\tau) y(t) = \mathcal{Z}^{-1}\left[\Phi_y(z)\right](\tau)$$

3. Per la soluzione di questo punto si vedano gli appunti del corso

Domanda 2.[8 Punti]

1. La varianza di $y(t)$ si calcola immediatamente per sostituzione:
$$\begin{aligned} Var\{y(t)\} &= \mathbb{E}y^2(t) = \mathbb{E}\left[(b_1 u(t-1) + b_2 u(t-2) + e_0(t))^2\right] \\ &= \mathbb{E}\left[(b_1 v(t-1) + (b_1+b_2) v(t-2) + b_2 v(t-3) + e_0(t))^2\right] \\ &= b_1^2 \sigma_v^2 + (b_1+b_2)^2 \sigma_v^2 + b_2^2 \sigma_v^2 + \sigma^2 \end{aligned}$$
dove nell'ultimo passaggio abbiamo usato il fatto che $v(t)$ è bianco e completamente scorrelato da $e(t)$; $\sigma_v^2 = 1$ è la varianza di $v(t)$.

2. Per stimare il parametro θ nel modello

$$y(t) = \theta u(t-1) + e(t)$$

basta calcolare il predittore in funzione di θ

$$y_\theta(t|t-1) = \theta u(t-1)$$

ed usare il criterio PEM

$$\hat{\theta}_N := \arg\min_\theta \frac{1}{N} \sum_{t=2}^{N} \left(y(t) - y_\theta(t|t-1)\right)^2$$

Per il nostro caso specifico lo stimatore si calcola in forma chiusa (il funzionale è quadratico in θ) e si ottiene lo stimatore dei minimi quadrati

$$\hat{\theta}_N = \frac{1}{N-1} \sum_{t=2}^{N} u(t-1)y(t) \left[\frac{1}{N-1} \sum_{t=2}^{N} u^2(t-1)\right]^{-1}$$

3. Sotto le ipotesi date la coppia $y(t), u(t)$ è ergodica e quindi lo stimatore PEM converge all'insieme dei punti di minimo del funzionale

$$\begin{aligned} J(\theta) &= \mathbb{E}\left(y(t) - y_\theta(t|t-1)\right)^2 \\ &= \mathbb{E}\left[\left((b_1-\theta)u(t-1) + b_2 u(t-2) + e_0(t)\right)^2\right] \\ &= \mathbb{E}\left[\left((b_1-\theta)v(t-1) + (b_1+b_2-\theta)v(t-2) + b_2 v(t-3) + e_0(t)\right)^2\right] \\ &= (b_1-\theta)^2 + (b_1+b_2-\theta)^2 + b_2^2 + \sigma^2 \end{aligned}$$

Per trovare (l'unico) punto di minimo $\theta_* = \arg\min_\theta J(\theta)$ basta calcolare la derivata in θ ed eguagliare a zero, ottenendo:

$$\theta_* = b_1 + \frac{b_2}{2}$$

Domanda 3.[8 Punti] Si indichi con E l'evento "decisione sbagliata" e con $\bar{E}$ il suo complementare, cioè l'evento "decisione corretta".

1. Usando il teorema della probabilità totale:

$$\begin{aligned} P[E] &= P[E|A]P[A] + P[E|B]P[B] \\ &= P[c(y_1,..,y_k) = B|A]P[A] + P[c(y_1,..,y_k) = A|B]P[B] \\ &= (1-p_A)\tfrac{2}{3} + (1-p_B)\tfrac{1}{3} \end{aligned}$$

dove si è usato che la probailità $P[A]$ che l'immagine estratta a caso dal database contenga un pallone da calcio è $\frac{2}{3}$ (perche due terzi delle immagini contengono palloni da calcio) e ovviamente $P[B] = 1 - P[A]$.

2. Utilizzando la regola di Bayes

$$\begin{aligned} P[A|c(y_1,..,y_k) = A] &= \tfrac{P[c(y_1,..,y_k)=A|A]P[A]}{P[c(y_1,..,y_k)=A]} \\ &= \tfrac{P[c(y_1,..,y_k)=A|A]P[A]}{P[c(y_1,..,y_k)=A|A]P[A]+P[c(y_1,..,y_k)=A|B]P[B]} \\ &= \tfrac{p_A\frac{1}{3}}{p_A\frac{2}{3}+(1-p_B)\frac{1}{3}} \end{aligned}$$

3. Per quanto riguarda la stima delle probabilità p_A e p_B basta operare come segue: si dividono le immagini disponibili in due gruppi. Il gruppo G_A contenente[4] $N_A = \lfloor \frac{1}{3}1000 \rfloor$ immagini ed il gruppo G_B contenente $N_B = 1000 - N_A$ immagini. Sia $\hat{C}_{i,G_A}$ l'esito del classificatore applicato all'i-esima immagine del gruppo A e sia $\hat{C}_{i,G_B}$ l'esito sulla i-esima immagine del gruppo G_B. Le probabilità p_A e p_B si possono stimare utilizzando le frequenze relative:
$$\hat{p}_A = f_{A|A} := \frac{1}{N_A} \#\{\hat{C}_{i,G_A} = A\}$$
$$\hat{p}_B = f_{B|B} := \frac{1}{N_B} \#\{\hat{C}_{i,G_B} = B\}$$

Domanda 4.[8 Punti]

1. Per calcolare la covarianza della coppia $x(t)$ e $y(t)$, $t \in [0, T]$ basta osservare che:
$$x(t) = x(0) + \sum_{k=0}^{t-1} r_x(k) \quad y(t) = y(0) + \sum_{k=0}^{t-1} r_y(k)$$
e ricordare che $x(0) = y(0) = 0$ e che $r_x(t)$ e $r_y(t)$ sono a media nulla, bianchi e indipendenti. Di conseguenza di ottiene che:
$$\mathbb{E}x(t) = \mathbb{E}y(t) = 0 \quad \forall t \in [0, T]$$
In aggiunta
$$cov(y(t), x(s)) = \mathbb{E}y(t)x(s) = \mathbb{E}\left[\sum_{k=0}^{t-1} r_y(k) \sum_{h=0}^{s-1} r_x(h)\right] = \sum_{h=0}^{s-1}\sum_{k=0}^{t-1} \mathbb{E}r_y(k)r_x(h) = 0$$
Manca quindi solo da calcolare $cov(x(t+\tau), x(t)) = \mathbb{E}x(t+\tau)x(t)$ e $cov(y(t+\tau)y(t)) = \mathbb{E}y(t+\tau)y(t)$. Applicando la definizione si ottiene, per $\tau \geq 0$,
$$\begin{aligned} cov(x(t+\tau), x(t)) &= \mathbb{E}x(t+\tau), x(t) \\ &= \mathbb{E}\left[\left(x(t) + \textstyle\sum_{k=0}^{\tau-1} r_x(t+k)\right) x(t)\right] \\ &= cov(x(t), x(t)) \end{aligned}$$
dove si ùsato il fatto che $r_x(t+k)$ è scorrelato da $x(t)$ per ogni $k \geq 0$. Infine
$$cov(x(t), x(t)) = \mathbb{E}x^2(t) = \mathbb{E}\left[\left(\sum_{k=0}^{t-1} r_y(k)\right)^2\right] = \sum_{k=0}^{t-1}\sum_{h=0}^{t-1} \mathbb{E}r_x(k)r_x(h) = t\sigma_r^2$$
dove nell'ultimo passaggio si è usato il fatto che $cov(r_x(k), r_x(h)) = \sigma_r^2\delta(k-h)$. Analogamente si può procedere per $y(t)$. In conclusione:
$$cov(x(t), x(s)) = \sigma_r^2 \min(t, s) \quad cov(y(t), y(s)) = \sigma_r^2 \min(t, s) \quad cov(y(t), x(s)) = 0$$

2. Per semplicità si assumerà in seguto che gli istanti a cui sono disponibili le misure siano del tipo $t_i = i$, $i = 0, .., T$. Per comodità di notazione si definisca:
$$Z_x^t = \begin{bmatrix} z_x(0) \\ z_x(1) \\ \vdots \\ z_x(t) \end{bmatrix} \quad Z_y^t = \begin{bmatrix} z_y(0) \\ z_y(1) \\ \vdots \\ z_y(t) \end{bmatrix}$$

[4] Poichè 1000/3 non è intero, assumimamo che in realtà ci siano $\lfloor \frac{1}{3}1000 \rfloor$ immagini con palloni da calcio.

Lo stimatore cercato ha quindi la forma[5]:

$$
\begin{aligned}
&\mathbb{E}\left[x(t)|z_x(s), z_y(s), s \in [0,T]\right] = \mathbb{E}\left[x(t)|Z_x^T, Z_y^T\right] = \\
&= \begin{bmatrix} cov(x(t), Z_x^T) & cov(x(t), Z_y^T) \end{bmatrix} \begin{bmatrix} cov(Z_x^T, Z_x^T) & cov(Z_x^T, Z_y^T) \\ cov(Z_y^T, Z_x^T) & cov(Z_y^T, Z_y^T) \end{bmatrix}^{-1} \begin{bmatrix} Z_x^T \\ Z_y^T \end{bmatrix} \\
&= \begin{bmatrix} cov(x(t), Z_x^T) & 0 \end{bmatrix} \begin{bmatrix} cov(Z_x^T, Z_x^T) & 0 \\ 0 & cov(Z_y^T, Z_y^T) \end{bmatrix}^{-1} \begin{bmatrix} Z_x^T \\ Z_y^T \end{bmatrix} \\
&= cov(x(t), Z_x^T) cov^{-1}(Z_x^T, Z_x^T) Z_x^T
\end{aligned}
$$

dove, usando il punto precedente:

$$cov(x(t), Z_x^T) = \sigma_t^2 \begin{bmatrix} 0 & 1 & 2 & \dots & t & t & \dots & t \end{bmatrix}$$

$$cov(Z_x^T, Z_x^T) = \begin{bmatrix} \sigma_n^2 & 0 & 0 & \dots & 0 & 0 \\ 0 & \sigma_n^2 + \sigma_r^2 & \sigma_r^2 & \dots & \sigma_r^2 & \sigma_r^2 \\ 0 & \sigma_r^2 & \sigma_n^2 + 2\sigma_r^2 & \dots & 2\sigma_r^2 & 2\sigma_r^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \sigma_r^2 & 2\sigma_r^2 & \dots & (T-1)\sigma_r^2 + \sigma_n^2 & (T-1)\sigma_r^2 \\ 0 & \sigma_r^2 & 2\sigma_r^2 & \dots & (T-1)\sigma_r^2 & \sigma_n^2 + T\sigma_r^2 \end{bmatrix}$$

Un'espressione analoga si ottiene per

$$\mathbb{E}\left[y(t)|z_x(s), z_y(s), s \in [0,T]\right]$$

3. Un algoritmo ricorsivo per il calcolo degli stimatori

$$\mathbb{E}\left[x(t)|z_x(s), z_y(s), s \in [0,t]\right] = \mathbb{E}\left[x(t)|Z_x^t, Z_y^t\right] = \hat{x}(t|t)$$

$$\mathbb{E}\left[y(t)|z_x(s), z_y(s), s \in [0,t]\right] = \mathbb{E}\left[x(t)|Z_x^t, Z_y^t\right] = \hat{y}(t|t)$$

si può semplicemente ricavare applicando il Filtro di Kalman al modello di stato che si ottiene dal testo dell'esercizio con stato

$$\xi(t) := \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

misure

$$z(t) := \begin{bmatrix} z_x(t) \\ z_y(t) \end{bmatrix}$$

con matrici di aggiornamento di stato e misura

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

e con le varianze dei rumori

$$Q = \sigma_r^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad R = \sigma_n^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad S = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

Il filtro va inizializzato con con condizioni iniziali

$$\hat{\xi}(0|-1) = [0\ 0]^\top \quad P(0|-1) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Ovviamente le stime cercate non sono altr che le stime "filtrate" dello stato, cioè:

$$\mathbb{E}\left[x(t)|z_x(s), z_y(s), s \in [0,t]\right] = [1\ 0]\hat{\xi}(t|t) \quad \mathbb{E}\left[y(t)|z_x(s), z_y(s), s \in [0,t]\right] = [0\ 1]\hat{\xi}(t|t)$$

[5] Si ricordi che al punto 1 si è visto che tutti i processi in gioco hanno media nulla.