Prosiding Seminar Nasional Integrasi Matematika dan Nilai Islami
Vol.3, No.1, September 2019, hlm. 346-351
p-ISSN: 2580-4596; e-ISSN: 2580-460X

Pendugaan Parameter Model Multivariat Normal Hidden Markov

Miftahul Fikri, Samsurizal
Jurusan Teknik Elektro, Sekolah Tinggi Teknik - PLN
miftahul@sttpln.ac.id, samsurizal@sttpln.ac.id

Info Artikel

Riwayat Artikel

Diterima: 21 Oktober 2019
Direvisi: 18 November 2019
Diterbitkan: 15 Januari 2020

Kata Kunci:

Model multivariat normal *hidden* Markov
Algoritme *Forward-Backward*
Algoritme *Expectation Maximization*

ABSTRAK

Model *hidden* Markov terdiri dari sepasang proses stokastik yaitu proses observasi dan proses yang mempengaruhi observasi. Proses stokastik yang mempengaruhi observasi ini diasumsikan membentuk rantai Markov dan tidak diamati. Model Multivariat Normal *hidden* Markov (model MNHM) adalah salah satu model *hidden* Markov dan proses observasinya jika diketahui proses yang mempengaruhinya diasumsikan menyebar multivariat Normal. Permasalahan utama model MNHM ialah menduga parameter yang memaksimumkan fungsi *likelihood*. Fungsi *likelihood* dihitung menggunakan algoritme *Forward-Backward*. Algoritme *Expectation Maximization* (algoritme EM) digunakan untuk memaksimumkan fungsi *likelihood*.

Korespondensi:

Miftahul Fikri,
Jurusan Teknik Elektro,
Sekolah Tinggi Teknik - PLN,
Jl. Lingkar Luar Duri Kosambi, Jakarta Barat, Indonesia 11750
miftahul@sttpln.ac.id

1. PENDAHULUAN

Terdapat banyak kejadian atau fenomena yang terjadi dalam kehidupan sehari-hari yang bersifat tidak pasti. Ketidakpastian ini dapat dijelaskan dengan proses stokastik. Hal ini dikarenakan proses stokastik merupakan suatu model yang dibangun dengan aturan-aturan peluang. Model ini dapat diterapkan dalam berbagai bidang pada kehidupan sehari-hari seperti nilai tukar rupiah, harga saham, kedatangan pelanggan ke suatu pusat layanan, dan banyaknya klaim pada suatu perusahaan asuransi.

Ketidakpastian pada suatu fenomena dapat disebabkan oleh beberapa faktor. Faktor-faktor penyebab ini seringkali sulit diamati. Model *hidden* Markov dapat diandalkan untuk memodelkan permasalahan ini.

Model *hidden* Markov terdiri dari sepasang proses stokastik, yaitu proses observasi dan proses yang mempengaruhinya (proses penyebab observasi). Proses stokastik penyebab observasi ini diasumsikan tidak diamati dan membentuk rantai Markov, yaitu peluang terjadinya penyebab kejadian pada suatu waktu tertentu hanya bergantung pada penyebab kejadian pada satu satuan waktu sebelumnya. Penyebab observasi ini biasa disebut *state*.

Berdasarkan kejadian observasinya, model *hidden* Markov dibedakan menjadi dua, yaitu model *hidden* Markov kontinu dan model *hidden* Markov diskrit. Model *hidden* Markov kontinu adalah model *hidden* Markov dengan kejadian observasinya kontinu, sedangkan model *hidden* Markov diskrit ialah model *hidden* Markov dengan kejadian observasinya diskrit. Model multivariat normal *hidden* Markov (model MNHM) adalah salah satu model *hidden* Markov kontinu di mana observasi (kejadian yang diamati) jika diketahui penyebab kejadiannya diasumsikan menyebar multivariate normal. Permasalahan utama pada model MNHM ialah menduga parameter yang memaksimumkan fungsi *likelihood* menggunakan algoritme *Expectation Maximization* (EM).

Jakarta Islamic Index (JII) adalah indeks saham syariah yang pertama kali diluncurkan di pasar modal Indonesia pada tanggal 3 Juli 2000. Konstituen JII hanya terdiri dari 30 saham syariah paling likuid yang tercatat di Bursa Efek Indonesia (BEI). Prediksi harga saham yang akurat merupakan hal yang sangat penting bagi para investor (terutama *trader*) karena salah satu keuntungan investor diperoleh dari selisih harga saham. JII merupakan suatu kejadian yang tidak dapat dipastikan kejadiannya sehingga dapat dimodelkan menggunakan model MNHM. Pada penelitian ini akan dilakukan pemodelan harga saham JII (*open*, *close*, *low*, *high*) menggunakan model multivariate normal *hidden* Markov.

2. METODE PENELITIAN

Model multivariate normal *hidden* Markov adalah model dengan waktu diskrit yang terdiri dari sepasang proses stokastik $\{X_t, Y_t\}_{t\in\mathbb{N}}$. $\{X_t\}_{t\in\mathbb{N}}$ merupakan penyebab kejadian yang diasumsikan tidak diamati dan membentuk suatu rantai Markov. Sedangkan $\{Y_t\}_{t\in\mathbb{N}}$ adalah proses observasinya yang hanya bergantung pada $\{X_t\}_{t\in\mathbb{N}}$. Kemudian peubah acak Y_t diketahui X_t diasumsikan menyebar multivariat normal, untuk setiap $t \in \mathbb{N}$, [1] dan [2].

Karakteristik Model Multivariat Normal *Hidden* Markov

1. Diasumsikan $\{X_t\}_{t\in\mathbb{N}}$ adalah rantai Markov diskret, homogen, tak tereduksi dan *ergodic* dengan ruang *state* $S_X = \{1,2,\cdots,m\}$.
2. Matriks peluang transisi $\Gamma = [\gamma_{ij}]$, di mana Γ matriks berukuran $m \times m$ dan $i, j \in S_X$, memenuhi:
   - $\gamma_{ij} = P(X_t = j|X_{t-1} = i) = P(X_2 = j|X_1 = i)$,
   - $\gamma_{ij} \geq 0$,
   - $\sum_{j=1}^{m} \gamma_{ij} = 1$, untuk setiap $i = 1,2,\dots,m$.

   Dalam model multivariat normal *hidden* Markov, Saat X_t berada pada *state* i $(i \in S_X)$, maka sebaran bersyarat Y_t jika diketahui $X_t = i$ $(t \in \mathbb{N})$ adalah peubah acak multivariat normal dengan parameter rataan $\boldsymbol{\mu}$ dan matriks kovarian $\boldsymbol{A}$. Untuk setiap $y \in \mathbb{R}^p$, matriks peluang dari proses observasi $\Pi = [\pi_{yi}]$, dengan

$$\pi_{yi} = P(Y_t = y|X_t = i) = \frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{\Sigma_i}|}}\, e^{\left(-\frac{(y-\boldsymbol{\mu_i})'\boldsymbol{\Sigma}_i^{-1}(y-\boldsymbol{\mu_i})}{2}\right)},$$

$$\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{\Sigma_i}|}}\, e^{\left(-\frac{(y-\boldsymbol{\mu_i})'\boldsymbol{\Sigma}_i^{-1}(y-\boldsymbol{\mu_i})}{2}\right)}\, dy_1\, dy_2 \dots dy_p = 1.$$

   [1]

3. Vektor peluang *state* awal $\delta = [\delta_i]$, di mana δ merupakan vektor berukuran $m \times 1$ dan $i \in S_X$, dengan

$$\delta_i = P(X_1 = i),$$

$$\sum_{i=1}^{m}\delta_i = 1.$$

   Karena rantai Markov $\{X_t\}_{t\in\mathbb{N}}$ diasumsikan rantai Markov yang *ergodic*, δ merupakan sebaran yang stasioner sehingga memenuhi

$$\Gamma\delta = \delta. \qquad (1)$$

4. Untuk setiap $t \in \mathbb{N}$ dan $y \in \mathbb{R}^n$, fungsi sebaran marginal dari Y_t, yaitu

$$P(Y_t = y) = \sum_{i=1}^{m} P(Y_t = y|X_t = i)P(X_t = i) = \sum_{i=1}^{m} \delta_i \pi_{yi} .$$

Berdasarkan pembahasan karakteristik model MNHM di atas, model multivariate normal *hidden* Markov $\{X_t, Y_t\}_{t\in\mathbb{N}}$ dicirikan oleh parameter $\phi = (\boldsymbol{\delta}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, dengan

$$\delta = [\delta_i] \quad i \in S_X,$$

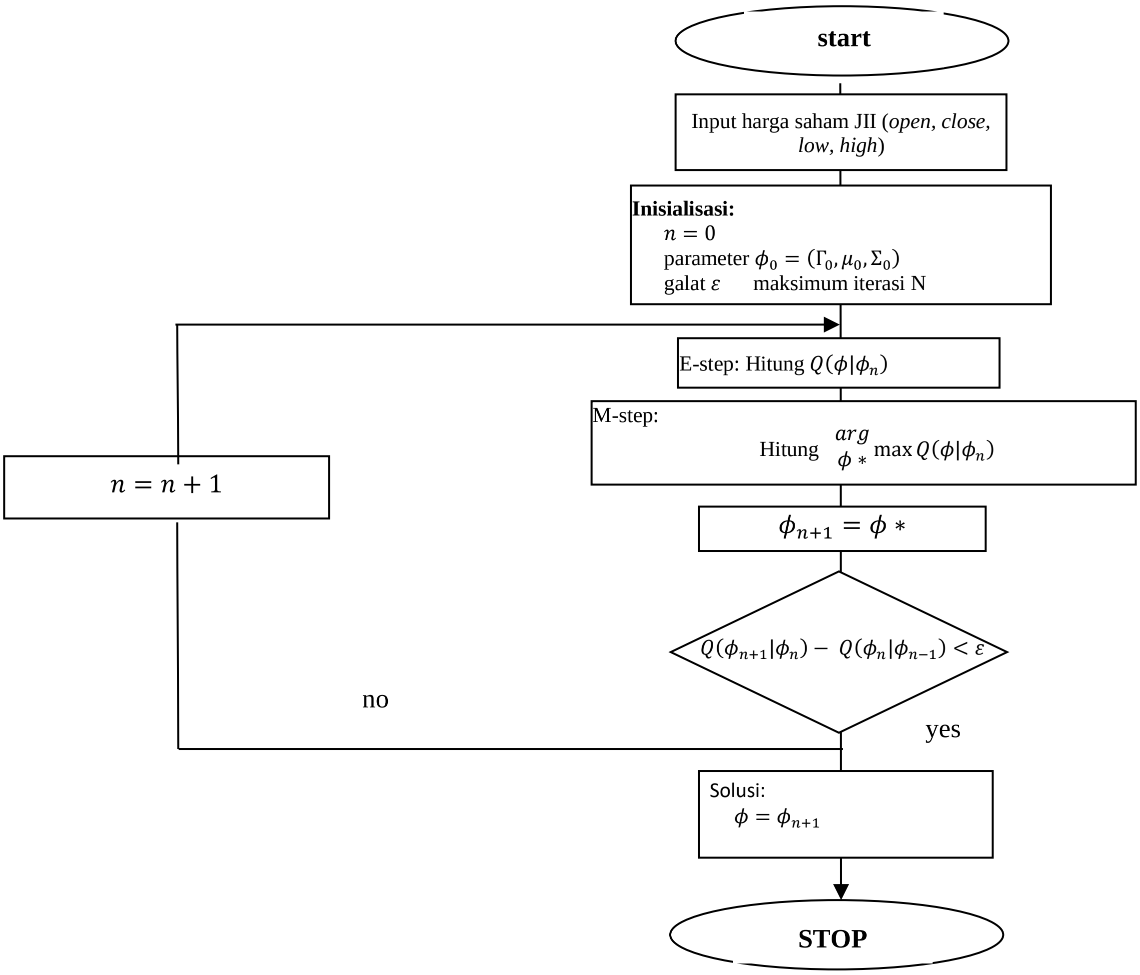
$$\Gamma = [\gamma_{ij}] \quad i, j \in S_X,$$

$$\mu = (\mu_1, \mu_2, \dots, \mu_m), \text{ dengan } \mu_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \vdots \\ \mu_{pi} \end{pmatrix}, \text{ untuk } i = 1,2, \dots, m.$$

$$\sum = (\Sigma_1, \Sigma_2, \dots, \Sigma_m), \text{ dengan } \Sigma_i = \begin{pmatrix} \sigma_{i11} & \sigma_{i12} & \cdots & \sigma_{i1p} \\ \sigma_{i21} & \sigma_{i22} & \cdots & \sigma_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ip1} & \sigma_{ip2} & \cdots & \sigma_{ipp} \end{pmatrix}, \text{ untuk } i = 1,2, \dots, m$$

Hal yang sangat penting pada model MNHM ialah mengestimasi parameter model. Akan tetapi berdasarkan persamaan (1), δ dengan mudah diperoleh ketika Γ diperoleh, sehingga cukup mengestimasi parameter $\phi = (\Gamma, \mu, \Sigma)$. Pada penelitian ini untuk mengestimasi parameter model dilakukan menggunakan metode maksimum *likelihood*. Setelah itu, model diterapkan pada harga saham JII saat dibuka (*open*), ditutup (*close*), terendah (*low*), tertinggi (*high*).

Langkah – langkah penelitian dibuat dalam diagram alir berikut:



Gambar 2.1. Diagram alir Penelitian

3. HASIL DAN PEMBAHASAN

Misalkan $y_1 = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{p1} \end{pmatrix}, y_2 = \begin{pmatrix} y_{12} \\ y_{22} \\ \vdots \\ y_{p2} \end{pmatrix}, . . . , y_T = \begin{pmatrix} y_{1T} \\ y_{2T} \\ \vdots \\ y_{pT} \end{pmatrix}$ merupakan p=4 buah data (open, high, low, closed) yang terjadi sepanjang T = 72 bulan.

Misalkan $\mathrm{M} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{p1} & \mu_{p2} & \cdots & \mu_{pm} \end{pmatrix}$

$\sum = (\Sigma_1, \Sigma_2, \dots, \Sigma_\mathrm{m})$, dengan $\Sigma_\mathrm{i} = \begin{pmatrix} \sigma_{i11} & \sigma_{i12} & \cdots & \sigma_{i1p} \\ \sigma_{i21} & \sigma_{i22} & \cdots & \sigma_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ip1} & \sigma_{ip2} & \cdots & \sigma_{ipp} \end{pmatrix}$, untuk $i = 1,2, \dots, m$

Misalkan

$$\delta_{i_1} = P(X_1 = i_1|\phi)$$

$$\pi_{y_t i_t} = P(Y_t = y_t|X_t = i_t, \phi) = \frac{1}{(2\pi)^{\frac{p}{2}} \left|\Sigma_{i_t}\right|^{\frac{1}{2}}} e^{-\frac{\left(y_t - \mu_{i_t}\right)' \Sigma_{i_t}^{-1} \left(y_t - \mu_{i_t}\right)}{2}},\ \text{untuk } i_t = 1,2, \dots, m \text{ dan } t = 1,2, \dots, T.$$

Fungsi *likelihood* dari proses observasi Y didefinisikan sebagai berikut:

$$\begin{aligned} L_T(\phi) &= P(Y_1 = y_1, Y_2 = y_2, \dots, Y_T = y_T|\phi) \\ &= p(y_1, y_2, \dots, y_T|\phi) \\ &= p(y|\phi) \\ &= \sum_{i_1=1}^{m} \dots \sum_{i_T=1}^{m} \left(\pi_{y_1 i_1} \pi_{y_2 i_2} \dots \pi_{y_T i_T}\right) \times \left(\delta_{i_1} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \dots \gamma_{i_{T-1} i_T}\right) \\ &= \sum_{i_1=1}^{m} \dots \sum_{i_T=1}^{m} \delta_{i_1} \pi_{y_1 i_1} \prod_{t=2}^{T} \gamma_{i_{t-1} i_t} \pi_{y_t i_t}. \end{aligned} \quad (2)$$

Permasalahan utama pada MPHM ialah mencari parameter $\phi^* \in \mathbf{\Phi}$ yang memaksimumkan fungsi *likelihood* $L_T(\phi)$. Untuk banyaknya data observasi T yang cukup besar, menghitung fungsi *likelihood* ini dibutuhkan waktu yang cukup lama. Algoritme *forward-backward* dapat digunakan untuk menangani masalah ini.

3.1. Pendugaan Parameter Rataan

Jika $\dfrac{\partial Q\left(\phi|\phi^{(k)}\right)}{\partial \mu_{uv}(\phi)} = 0$, maka akan diperoleh

$$\mu_{uv}\left(\phi^{(k+1)}\right) = \frac{\sum_{t=1}^{T} \alpha_t\left(v|\phi^{(k)}\right) \beta_t\left(v|\phi^{(k)}\right) \left(2\, s_{vuu} y_{ut} + \sum_{\substack{k=1 \\ k \neq u}}^{p} s_{vku} \left(y_{kt} - \mu_{kv}\right) + \sum_{\substack{k=1 \\ k \neq u}}^{p} s_{vuk} \left(y_{kt} - \mu_{kv}\right)\right)}{2\, s_{vuu} \sum_{t=1}^{T} \alpha_t\left(v|\phi^{(k)}\right) \beta_t\left(v|\phi^{(k)}\right)},$$

untuk $u = 1,2, \dots, m$, $v = 1,2, \dots, p$, dan

$$\Sigma_{\mathrm{i}}^{-1} = \begin{pmatrix} \mathrm{s_{i11}} & \mathrm{s_{i12}} & \cdots & \mathrm{s_{i1p}} \\ \mathrm{s_{i21}} & \mathrm{s_{i22}} & \cdots & \mathrm{s_{i2p}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{s_{ip1}} & \mathrm{s_{ip2}} & \cdots & \mathrm{s_{ipp}} \end{pmatrix}, \text{ untuk } i = 1,2,\dots,m.$$

3.2 Pendugaan Parameter Simpangan Baku

Jika $\dfrac{\partial Q\left(\phi|\phi^{(k)}\right)}{\partial \sigma_{uvw}(\phi)} = 0$, maka akan diperoleh

$$\sigma_{\mathrm{uvw}} = \frac{\left(\sum_{\substack{j=1 \\ j\neq w}}^{p} (-1)^{v+j}\sigma_{uvj}\Sigma_{uvj}\right)\sum_{t=1}^{T}\alpha_t\left(u|\phi^{(k)}\right)\beta_t\left(u|\phi^{(k)}\right)\left(\left((-1)^{v+w}\Sigma_{uvw}\right) + (\mathrm{y_t}-\mu_{\mathrm{u}})'\,A\,(\mathrm{y_t}-\mu_{\mathrm{u}})\right) - \sum_{t=1}^{T}\alpha_t\left(u|\phi^{(k)}\right)\beta_t\left(u|\phi^{(k)}\right)\left((\mathrm{y_t}-\mu_{\mathrm{u}})'\,B\,(-1)^{v+w}\Sigma_{uvw}\,(\mathrm{y_t}-\mu_{\mathrm{u}})\right)}{-(\Sigma_{uvw})^2\sum_{t=1}^{T}\alpha_t\left(u|\phi^{(k)}\right)\beta_t\left(u|\phi^{(k)}\right)},$$

untuk $u = 1,2,\dots,m;\quad v,w = 1,2,\dots,p.$

dengan

A

$$= \begin{pmatrix} (-1)^{1+1}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{u11v^*w^*}} & (-1)^{2+1}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{u21v^*w^*}} & \cdots & 0 & \cdots & (-1)^{\mathrm{p}+1}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{up1v^*w^*}} \\ (-1)^{1+2}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{u12v^*w^*}} & (-1)^{2+2}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{u22v^*w^*}} & \cdots & 0 & \cdots & (-1)^{\mathrm{p}+2}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{up2v^*w^*}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ (-1)^{1+\mathrm{p}}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{u1pv^*w^*}} & (-1)^{2+\mathrm{p}}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{u2pv^*w^*}} & \cdots & 0 & \cdots & (-1)^{\mathrm{p+p}}(-1)^{\mathrm{v^*+w^*}}\Sigma_{\mathrm{uppv^*w^*}} \end{pmatrix},$$

$$B = \begin{pmatrix} x_{11} & x_{12} & \cdots & - & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & - & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ - & - & \cdots & - & \cdots & - \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & - & \cdots & x_{pp} \end{pmatrix},$$

di mana

$$x_{ij} = (-1)^{j+i}\left(\sum_{\substack{k=1 \\ k\neq i;k\neq w}}^{p} (-1)^{v^*+k^*}\sigma_{\mathrm{uvk}}\Sigma_{\mathrm{ujiv^*k^*}}\right),$$

$$v^* = \begin{cases} v-1, & j < v \\ 0, & j = v \\ v, & j > v \end{cases}$$

$$w^* = \begin{cases} w-1, & i < w \\ 0, & i = w \\ w, & i > w \end{cases}$$

Σ_{ijklm}: Determinan dari suatu bentuk matriks dengan cara (menghapus baris j dan kolom k dari matriks Σ_{i}, kemudian hapus baris l dan kolom m).

3.3. Pendugaan Parameter Peluang Transisi

Misalkan $G\left(\phi|\phi^{(k)}\right) = Q\left(\phi|\phi^{(k)}\right) - \sum_{i=1}^{m}\theta_i\left(\sum_{j\in S_X}\gamma_{ij}(\phi) - 1\right)$.

jika $\frac{\partial G(\phi|\phi^{(k)})}{\partial \gamma_{ij}(\phi)} = 0,$ maka akan diperoleh

$$\gamma_{ij}\left(\phi^{(k+1)}\right) = \frac{\sum_{t=1}^{T-1} \gamma_{ij}\left(\phi^{(k)}\right)\alpha_t\left(i|\phi^{(k)}\right)P\left(Y_{t+1} = y_{t+1}|X_{t+1} = j, \phi^{(k)}\right)\beta_{t+1}\left(j|\phi^{(k)}\right)}{\sum_{t=1}^{T-1} \alpha_t\left(i|\phi^{(k)}\right)\beta_t\left(i|\phi^{(k)}\right)}.$$

4. KESIMPULAN

Kesimpulan

Didapatkan rumus penduga parameter model multivariat normal *hidden* Markov menggunakan *algoritme Expectation Maximization*

Saran

Untuk penelitian selanjutnya, pemodelan harga saham JII dapat menggunakan model normal hidden markov tak homogen.

5. UCAPAN TERIMAKASIH

1 Terimakasih kami ucapkan kepada STT-PLN yang telah mendukung pelaksanaan penelitian ini.
2. Terimakasih kami ucapkan kepada Kemenristek Dikti yang yang telah mensupport dana pada penelitian ini

DAFTAR PUSTAKA

[1] Spezia L. 2010. Bayesian analysis of multivariate Gaussian hidden Markov models with an unknown number of regimes. Journal of Time Series Analysis. 31: 1-1

[2] Paroli R, Spezia L. 1999. Gaussian hidden Markov models: parameters estimation and applications to air pollution data. [E.P.n. 94]. Milan (IT): Universita Cattolica Del Sacro Cuore.