



TRATAMIENTO DIGITAL DE SEÑALES

Ingeniería de Telecomunicación (4º, 2º c)

Unidad 5ª: El algoritmo EM. Mezclas de gaussianas

Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

Un problema de estimación ML

E: Se dispone de las observaciones $\{x^{(k)}\}_{k=1}^K$, tomadas independientemente una de otra, de una mezcla de gaussianas

$$p(x) = \rho \frac{1}{\sqrt{2\pi v_1}} \exp\left[-\frac{(x-m_1)^2}{2v_1}\right] + (1-\rho) \frac{1}{\sqrt{2\pi v_2}} \exp\left[-\frac{(x-m_2)^2}{2v_2}\right] = \rho p_1(x) + (1-\rho)p_2(x)$$

siendo los parámetros $s = [\rho m_1 v_1 m_2 v_2]^T$ valores deterministas desconocidos.

Establezca las ecuaciones ML para la estimación de dichos parámetros.

Discuta la dificultad de resolución de dichas ecuaciones.

$$p(\{x\}^{(k)} / s) = \prod_{k=1}^K p(x^{(k)} / s) = \prod_{k=1}^K \left\{ \rho p_1(x^{(k)} / m_1, v_1) + (1-\rho) p_2(x^{(k)} / m_2, v_2) \right\}$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

y las ecuaciones son $\frac{\partial \ln}{\partial s_i} = 0$; es decir

$$\frac{\partial \ln}{\partial \rho} = \sum_{k=1}^K \frac{p_1(x^{(k)} | m_1, v_1) - p_2(x^{(k)} | m_2, v_2)}{p(x^{(k)} | s)} = 0$$

$$\frac{\partial \ln}{\partial m_1} = \sum_{k=1}^K \frac{p_1(x^{(k)} | m_1, v_1)}{p(x^{(k)} | s)} (-2) \frac{x^{(k)} - m_1}{v_1} = 0$$

$$\frac{\partial \ln}{\partial v_1} = \sum_{k=1}^K \frac{1}{p(x^{(k)} | s)} \rho \frac{1}{\sqrt{2\pi}} \left[-\frac{1}{2} v_1^{-3/2} + \frac{(x^{(k)} - m_1)^2}{2v_1^2} \right] \exp \left[-\frac{(x^{(k)} - m_1)^2}{2v_1} \right] = 0$$

y análogamente para m_2 y v_2 .

Es un sistema de ecuaciones no lineales: se necesita utilizar algún método de búsqueda para su solución, y al acoplar entre ecuaciones hace series

Cartagena99

CLASES PARTICULARES, TUTORIAS TECNICAS ONLINE
LLAMA O ENVIA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4.IT/ATSC-DTC/UCIIM

E: Considere el mismo caso anterior, pero incluyendo con cada muestra la observación de una variable indicadora que señale si $x^{(k)}$ ha sido generada por la primera gaussiana o por la segunda ($z^{(k)}$: $[1, 0]^T$ si p_1 , $[0, 1]^T$ si p_2).

Indíquese cómo se procedería en esta situación.

Se dividiría el conjunto de muestras en $\{x^{(k)}\}_{k=1}^{K'}$ correspondientes a la primera gaussiana y $\{x^{(k)}\}_{k=K'+1}^K$ correspondientes a la segunda; con cada uno de ellos se estimaría m_1 , v_1 y m_2 , v_2 , respectivamente, en la forma ya conocida: estimadores muestrales.

Quedaría entonces aplicar la primera ecuación anterior para la estimación de ρ :

$$\sum_{k=1}^K \frac{p_1(x^{(k)} | \hat{m}_1, \hat{v}_1) - p_2(x^{(k)} | \hat{m}_2, \hat{v}_2)}{\rho p_1(x^{(k)} | \hat{m}_1, \hat{v}_1) + (1 - \rho)p_2(x^{(k)} | \hat{m}_2, \hat{v}_2)} = 0$$

$$\sum_{k=1}^K \frac{1}{\rho + \frac{p_2(x^{(k)} | \hat{m}_2, \hat{v}_2)}{p_1(x^{(k)} | \hat{m}_1, \hat{v}_1) - p_2(x^{(k)} | \hat{m}_2, \hat{v}_2)}} = 0$$

$$p_1(x^{(k)} | \hat{m}_1, \hat{v}_1) - p_2(x^{(k)} | \hat{m}_2, \hat{v}_2)$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4 IT/ATSC-DTC/UCIIM

Discusión

Se ha visto que, en una situación moderadamente compleja, la resolución de las ecuaciones ML puede resultar complicada. Sin embargo, si se hubiese contado con la observación del indicador, las cosas serían elementales.

Lo que quiere decir que, si se tuviese acceso a observar una variable $y = (x, z)$, se podría proceder sin dificultad. Pero no si tenemos observaciones “incompletas”, de x (la variable **observable**).

Ante situaciones de este tipo se abre la posibilidad de construir artificialmente una variable “completa” y ; ganando ventaja computacional si sabemos qué hacer para, en definitiva, maximizar $p(x|s)$ (que es el problema ML planteado).

A ello nos ayuda el algoritmo que se expone acto seguido.

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TDS/4-IT/ATSC-DTC/UCIIM

El Algoritmo EM (“Expectation-Maximization”)

Dempster, Laird y Rubin propusieron proceder así: establecida la expresión

$$p(\mathbf{y} | \mathbf{s}) = p(\mathbf{y} | \mathbf{x}, \mathbf{s}) p(\mathbf{x} | \mathbf{s})$$

y sobre la forma

$$\ln p(\mathbf{y} | \mathbf{s}) = \ln p(\mathbf{y} | \mathbf{x}, \mathbf{s}) + \ln p(\mathbf{x} | \mathbf{s})$$

* inicialícese \mathbf{s} en $\mathbf{s}(1)$

* en el paso n , aplíquese

- una **etapa E**: promediando $\ln p(\mathbf{y} | \mathbf{s})$ respecto a $p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n))$; es decir, calculando

$$\int \ln p(\mathbf{y} | \mathbf{x}, \mathbf{s}) p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n)) d\mathbf{y} + \ln p(\mathbf{x} | \mathbf{s})$$

que indicaremos como

$$Q(\mathbf{s}, \mathbf{s}(n)) = E_{|\mathbf{s}(n)} \{ \ln p(\mathbf{y} | \mathbf{s}) \}$$

- una **etapa M**: $\mathbf{s}(n+1) = \arg \{ \max_{\mathbf{s}} Q(\mathbf{s}, \mathbf{s}(n)) \}$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TDS/4.IT/ATSC-DTC/UCIIM

En el apéndice se prueba que este algoritmo maximiza $p(\mathbf{x} | \mathbf{s})$

Notas

- El Algoritmo EM no está exento de padecer problemas de detención en máximos locales.
- Si se sustituye la etapa M (p. ej., por dificultad analítica) por un mero aumento de Q, el algoritmo resultante (**GEM**: “Generalized EM”) tiene igual sentido.
- Si en las expresiones anteriores se añade $\ln p(\mathbf{s})$ a $\ln p(\mathbf{y} | \mathbf{s})$, se tiene la formulación EM para planteamiento MAP.

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

Ejemplo de aplicación (Modelado EM-GM)

A: Formúlese el Algoritmo EM para la mezcla de dos Gaussianas, utilizando la variable indicadora z como no observable.

Se tiene como variable completa $y = zx$; cuya ddp será ρp_1 si z toma el valor $[1, 0]^T$, y $(1 - \rho)p_2$ si z es $[0, 1]^T$; es decir

$$p(\{y^{(k)}\} | s) = \prod_{k=1}^K z^{(k)T} \begin{bmatrix} \rho p_1(x^{(k)} | m_1, v_1) \\ (1 - \rho)p_2(x^{(k)} | m_2, v_2) \end{bmatrix}$$

$$\ln p(\{y^{(k)}\} | s) = \sum_{k=1}^K z^{(k)T} \begin{bmatrix} \ln \rho p_1(x^{(k)} | m_1, v_1) \\ \ln (1 - \rho)p_2(x^{(k)} | m_2, v_2) \end{bmatrix}$$

con lo que el Algoritmo EM resulta:

- etapa E

$$\rho = \frac{1}{K} \sum_{k=1}^K z^{(k)T} \begin{bmatrix} \ln \rho p_1(x^{(k)} | m_1, v_1) \\ \ln (1 - \rho)p_2(x^{(k)} | m_2, v_2) \end{bmatrix}$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

$$E_{/s(n)} \{ z_1^k \} = \frac{\rho(n) p_1(x^{(k)} / m_1(n), v_1(n))}{p(x^{(k)} / s(n))} = w^{(k)}(n)$$

$$E_{/s(n)} \{ z_2^{(k)} \} = \frac{(1 - \rho(n)) p_2(x^{(k)} / m_2(n), v_2(n))}{p(x^{(k)} / s(n))} = 1 - w^{(k)}(n)$$

con lo que resulta

$$Q(s, s(n)) = \sum_{k=1}^K [w^{(k)}(n) \quad 1 - w^{(k)}(n)] \left[\begin{array}{l} \ln \rho p_1(x^{(k)} / m_1, v_1) \\ \ln(1 - \rho) p_2(x^{(k)} / m_2, v_2) \end{array} \right]$$

- Etapa M

$$* \frac{\partial Q(s, s(n))}{\partial \rho} = \sum_{k=1}^K \left[\frac{w^{(k)}(n)}{\rho} - \frac{1 - w^{(k)}(n)}{1 - \rho} \right] = 0$$

$$\sum_{k=1}^K [(1 - \rho) w^{(k)}(n) - \rho (1 - w^{(k)}(n))] = 0 ;$$

$$\sum_{k=1}^K (w^{(k)}(n) - \rho) = 0$$

de donde

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

$$* \frac{\partial Q(s, s(n))}{\partial m_1} = \sum_{k=1}^K w^{(k)}(n) \frac{\partial \ln p_1(x^{(k)} / m_1, v_1)}{\partial m_1} = 0$$

$$\sum_{k=1}^K \frac{\rho(n) p_1(x^{(k)} / m_1(n), v_1(n))}{p(x^{(k)} / s(n))} \frac{x^{(k)} - m_1}{v_1} = 0$$

$$\sum_{k=1}^K \frac{p_1(x^{(k)} / m_1(n), v_1(n))}{p(x^{(k)} / s(n))} (x^{(k)} - m_1) = 0$$

de donde

$$m_1(n+1) = \frac{\sum_{k=1}^K \frac{p_1(x^{(k)} / m_1(n), v_1(n))}{p(x^{(k)} / s(n))} x^{(k)}}{\sum_{k=1}^K \frac{p_1(x^{(k)} / m_1(n), v_1(n))}{p(x^{(k)} / s(n))}}$$

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

Cartagena99

TDS/4. IT/ATSC-DTC/UCIIM

$$\begin{aligned}
 * \quad \frac{\partial Q(s, s(n))}{\partial v_1} &= \sum_{k=1}^K w^{(k)}(n) \frac{\partial \ln p_1(x^{(k)} | m_1, v_1)}{\partial v_1} = 0 \\
 \sum_{k=1}^K \frac{\rho(n) p_1(x^{(k)} | m_1(n), v_1(n))}{p(x^{(k)} | s(n))} &\left[-\frac{1}{2v_1} + \frac{1}{2v_1^2} (x^{(k)} - m_1)^2 \right] = 0 \\
 \sum_{k=1}^K \frac{p_1(x^{(k)} | m_1(n), v_1(n))}{p(x^{(k)} | s(n))} &\left[(x^{(k)} - m_1)^2 - v_1 \right] = 0
 \end{aligned}$$

de donde

$$v_1(n+1) = \frac{\sum_{k=1}^K \frac{p_1(x^{(k)} | m_1(n), v_1(n))}{p(x^{(k)} | s(n))} (x^{(k)} - m_1(n+1))^2}{\sum_{k=1}^K \frac{p_1(x^{(k)} | m_1(n), v_1(n))}{p(x^{(k)} | s(n))}}$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

Notando que

$$\sum_{k=1}^K \frac{\rho(n) p_1(x^{(k)} / m_1(n), v_1(n))}{p(x^{(k)} / s(n))} = \sum_{k=1}^K w^{(k)}(n)$$

y lo mismo con subíndices 2 y $1-w^{(k)}(n)$, son posibles expresiones más compactas:

$$\rho(n+1) = \frac{1}{K} \sum_{k=1}^K w^{(k)}(n)$$

$$m_1(n+1) = \frac{\sum_{k=1}^K w^{(k)}(n) x^{(k)}}{\sum_{k=1}^K w^{(k)}(n)}$$

$$v_1(n+1) = \frac{\sum_{k=1}^K w^{(k)}(n) (x^{(k)} - m_1(n+1))^2}{\sum_{k=1}^K w^{(k)}(n)}$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

La generalización de lo anterior a una mezcla de G gaussianas multidimensionales (**GM**, “Gaussian Mixture”) es inmediata: si la ddp es

$$p(\mathbf{x} | \mathbf{s}) = \sum_{j=1}^G \rho_j p_j(\mathbf{x} | \mathbf{m}_j, \mathbf{V}_j)$$

se tiene

$$\rho_j(n+1) = \frac{1}{K} \sum_{k=1}^K w_j^{(k)}(n)$$

$$\mathbf{m}_j(n+1) = \frac{\sum_{k=1}^K w_j^{(k)}(n) \mathbf{x}^{(k)}}{\sum_{k=1}^K w_j^{(k)}(n)}$$

$$\mathbf{V}_j(n+1) = \frac{\sum_{k=1}^K w_j^{(k)}(n) (\mathbf{x}^{(k)} - \mathbf{m}_j(n+1)) (\mathbf{x}^{(k)} - \mathbf{m}_j(n+1))^T}{\sum_{k=1}^K w_j^{(k)}(n)}$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM

Apéndice: convergencia del Algoritmo EM

Se tiene que

$$\begin{aligned} Q(\mathbf{s}, \mathbf{s}(n)) &= E_{|\mathbf{s}(n)} \{ \ln p(\mathbf{y} | \mathbf{x}, \mathbf{s}) \} + \ln p(\mathbf{x} | \mathbf{s}) \\ &= H(\mathbf{s}, \mathbf{s}(n)) + \ln p(\mathbf{x} | \mathbf{s}) \end{aligned}$$

y es

$$\begin{aligned} H(\mathbf{s}, \mathbf{s}(n)) - H(\mathbf{s}(n), \mathbf{s}(n)) &= E_{|\mathbf{s}(n)} \left\{ \ln \frac{p(\mathbf{y} | \mathbf{x}, \mathbf{s})}{p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n))} \right\} = \\ &= \int_{\mathbf{Y}} \ln \left[\frac{p(\mathbf{y} | \mathbf{x}, \mathbf{s})}{p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n))} \right] p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n)) d\mathbf{y} \leq 0 \quad (\text{si y sólo si } \mathbf{s} = \mathbf{s}(n)) \end{aligned}$$

debido a la versión límite de la desigualdad de Jensen: para una función convexa f y unos pesos $p_j \geq 0$ tales que $\sum p_j = 1$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

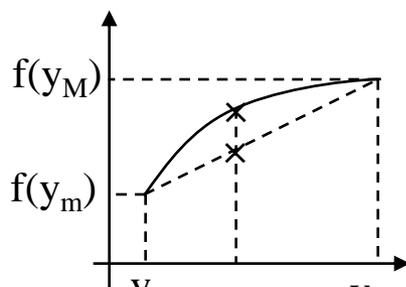
TDS/4-IT/ATSC-DTC/UCIIM

tomando \ln como f y $p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n))$ en el papel de las p_j

$$\begin{aligned} H(\mathbf{s}, \mathbf{s}(n)) - H(\mathbf{s}(n), \mathbf{s}(n)) &\leq \ln \int_{\mathbf{Y}} \left[\frac{p(\mathbf{y} | \mathbf{x}, \mathbf{s})}{p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n))} \right] p(\mathbf{y} | \mathbf{x}, \mathbf{s}(n)) d\mathbf{y} = \\ &= \ln \int_{\mathbf{Y}} p(\mathbf{y} | \mathbf{x}, \mathbf{s}) d\mathbf{y} \end{aligned}$$

de modo que en cada paso M , al decrecer H , se incrementa $p(\mathbf{x} | \mathbf{s})$.

La prueba de la desigualdad de Jensen es elemental: f es convexa si, con $0 \leq p \leq 1$



$$f [py_m + (1-p) y_M] \geq pf(y_m) + (1-p)f(y_M)$$

ya que $py_m + (1-p) y_M$ es un punto del intervalo $[y_m, y_M]$

CLASES PARTICULARES, TUTORIAS TÉCNICAS ONLINE
LLAMA O ENVIA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

Cartagena99

TDS/4-IT/ATSC-DTC/UCIIM

Probada la desigualdad para dos sumandos, se supone cumplida para J ; como para $J+1$ se tiene

$$p_{J+1}y_{J+1} + \sum_{j=1}^J p_j y_j = p_{J+1}y_{J+1} + (1-p_{J+1}) \sum_{j=1}^J \frac{p_j}{1-p_{J+1}} y_j = p_{J+1}y_{J+1} + (1-p_{J+1}) \sum_{j=1}^J p'_j y_j$$

con $p'_j \geq 0$, $\sum_{j=1}^J p'_j = 1$; aplicando la convexidad

$$f\left(\sum_{j=1}^{J+1} p_j y_j\right) \geq p_{J+1} f(y_{J+1}) + (1-p_{J+1}) f\left(\sum_{j=1}^J p'_j y_j\right)$$

y por la desigualdad para J

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

TDS/4.IT/ATSC-DTC/UCIIM

que es, en definitiva

$$f\left(\sum_{j=1}^{J+1} p_j y_j\right) \geq \sum_{j=1}^{J+1} p_j f(y_j)$$

y cuya versión límite es

$$f\left(\int y(x)p(x)dx\right) \geq \int f[y(x)p(x)]dx$$

con $p(x) \geq 0$ e $\int p(x)dx = 1$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TDS/4_IT/ATSC-DTC/UCIIM