



TRATAMIENTO DIGITAL DE SEÑALES

Ingeniería de Telecomunicación (4º, 2º c)

Unidad 8ª: Estimación de Densidades de Probabilidad

Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

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



de tipos de procedimientos

Paramétricos: en los que se acepta una forma analítica para la ddp (p.ej., gaussiana) y se estiman sus parámetros (media y varianza, en el ejemplo)

Procedimientos sencillos, pero poco robustos.

Paramétricos: en los que la ddp se modela según una expresión con suficiente capacidad de aproximación (habitualmente una mezcla: p. ej. , de gaussianas), y se estiman sus parámetros (p.ej., vía Algoritmo EM).

C/DTC/UCIIM

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

Son procedimientos eficaces, pero no exentos de dificultades; como:

- determinación del “orden”
- convergencia

métricos: no se realiza asunción alguna sobre la forma de la ddp, y se recurre a estimarla a partir de la aplicación de propiedades generales de la ddp.

Procedimientos robustos: pero también con sus inconvenientes, como:

- alta necesidad de memoria en diseño
- alta carga computacional en aplicación

C/DTC/UCIIM



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

Transición a los procedimientos no paramétricos

El procedimiento clásico por antonomasia, en situaciones unidimensionales (para visualizar el resultado), es el **histograma**: se divide el espacio muestral en segmentos o cuadrados, normalmente de dimensión fija, y se asigna a cada uno una altura uniforme igual al número de muestras que contiene, k (proporciona así la forma de la ddp).

Inconvenientes:

1. La aproximación es escalonada

2. Las regiones básicas utilizadas (segmentos, cuadrados, cubos, ... hipercubos) son inadecuadas en muchas dimensiones: un hipercubo, al aumentar las dimensiones, reduce mucho su volumen respecto a su lado, y presenta un aspecto de “puercoespín”; lo que es, obviamente, inapropiado para representar los datos (puede incluir datos muy distantes y separar datos muy próximos)

UCIIM/DTC/C



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

Es importante, en cualquier caso, reflexionar sobre la teoría subyacente aplicación de los histogramas: en definitiva, toman una región y le n el número de muestras que contiene. Lo que se entiende considerando n general, la probabilidad de encontrarse en una región R de una ddp es

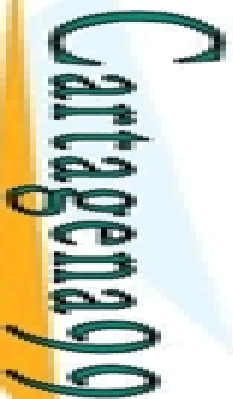
$$P = \int_R p(\mathbf{x}') \, d\mathbf{x}'$$

De acuerdo con la definición de valor medio de una integral definida, puede expresarse

$$P = p(\mathbf{x}) V_R$$

donde V_R es el volumen de R y \mathbf{x} un cierto punto interior a R

.../DTC/UCIIM



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

Aceptando que P se puede aproximar por la “frecuencia relativa” de las ras en R , resulta

$$p(\mathbf{x})V_R \approx k / K$$

→ k el número de muestras en R y K el número total de muestras; de donde

$$p(\mathbf{x}) \approx \frac{k}{KV_R}$$

→ $p(\mathbf{x})$ la que tomaría el histograma si se fuerza área unitaria para $p(\mathbf{x})$.

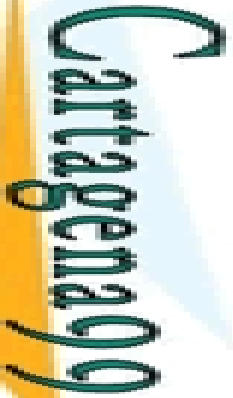
Los métodos particulares dependen de la elección de R .

C/DTC/UCHIM



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



Debe resaltarse que se establece un compromiso:

produce un efecto de promediado espacial, al tomar x como un valor medio integral: por lo que conviene emplear regiones R pequeñas;

pero, al mismo tiempo, para que la aproximación de P sea razonable, ha de mantenerse k/K lejos de 0: lo que supone que R no puede hacerse definitivamente pequeña;

por lo que

conviene que K sea alto: lo que origina cargas elevadas;

conviene elegir las regiones para mantener k razonablemente alto: es decir, en función de la localización de los datos (a diferencia de los histogramas)

C/DTC/UCIIM

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

cedimiento de las Ventanas (Núcleos) de Parzen

es la función indicadora de un hipercubo de lado 1 centrado en el origen

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } |\mathbf{x}_n| < 1/2, \\ 0, & \text{caso contrario} \end{cases} \quad \forall n \text{ (componentes)}$$

que

$$\phi\left(\frac{\mathbf{x} - \mathbf{x}^{(k')}}{h}\right) = \begin{cases} 1, & \text{si } |\mathbf{x}_n - \mathbf{x}_n^{(k')}| < h/2, \\ 0, & \text{caso contrario} \end{cases} \quad \forall n$$

que, dado \mathbf{x} , cuenta 1 si $\mathbf{x}^{(k')}$ está en el hipercubo, y 0 si no; así:

$$k = \sum_{k'=1}^K \phi\left(\frac{\mathbf{x} - \mathbf{x}^{(k')}}{h}\right)$$

se tiene el estimador

$$\hat{p}(\mathbf{x}) = \frac{1}{Kh^N} \sum_{k'=1}^K \phi\left(\frac{\mathbf{x} - \mathbf{x}^{(k')}}{h}\right)$$

(para el hipercubo)

C/DTC/UCHIM

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

Esta forma de actuar soluciona el problema de los histogramas cuando se maneja con prefijar regiones, ya que el hipercubo se centra en el \mathbf{x} observado; pero no las malas características de los hipercubos. Por ello, se maneja la forma de éstos, manejando **ventanas** o **núcleos** más “lisos”, con las siguientes condiciones de

o no negatividad: $\phi(\mathbf{x}) \geq 0$

o volumen unitario: $\int \phi(\mathbf{x}) d\mathbf{x} = 1$

Con estas condiciones se obtienen los **estimadores de Parzen**

$$\hat{p}(\mathbf{x}) = \frac{1}{Kh^N} \sum_{k'=1}^K \phi\left(\frac{\mathbf{x} - \mathbf{x}^{(k')}}{h}\right)$$

La elección de h establece el compromiso entre resolución y consistencia que se ha discutido).

UCI/UCIIM

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

Es muy frecuente utilizar ventanas gaussianas

$$\phi(\mathbf{x} - \mathbf{x}^{(k')}) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}^{(k')})^T \mathbf{V}^{-1} (\mathbf{x} - \mathbf{x}^{(k')})\right]$$

es \mathbf{V} lo que marca la concentración), y sobre todo sus versiones circulares

$$\phi(\mathbf{x} - \mathbf{x}^{(k')}) = \frac{1}{(2\pi v)^{N/2}} \exp\left[-\frac{1}{2v} \|\mathbf{x} - \mathbf{x}^{(k')}\|^2\right]$$

ventanas para este procedimiento.

No debe olvidarse que la elección de v marcará el comportamiento del
 vector: “picudo” si es pequeña, muy alisado si es grande.

C/DTC/UCIIM

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

Procedimiento de los **k Vecinos más Próximos** (k-NN) (“Nearest Neighbors”)

En este caso, V_R se fija en función de las posiciones de los datos: el volumen V_R (de una región R de forma razonable: hiperesférica, p.ej.) en torno a \mathbf{x} debe haber k muestras; y así

$$\hat{p}(\mathbf{x}) = \frac{k}{KV_R}$$

Nótese que es V_R lo que varía en esta expresión.

Hay métodos rápidos para seleccionar las k muestras más próximas.

T: Estúdiese un procedimiento rápido para aplicar k-NN.

C/DTC/UCIIM

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

Estimación no paramétrica de las probabilidades a posteriori

NN

Si suponemos una situación en que las muestras pertenezcan a clases C_c ,
 ma

$$\hat{p}(\mathbf{x}) = \sum_{c'=1}^c \hat{p}(\mathbf{x}, C_{c'}) = \frac{k}{KV_R}$$

la misma región, dividiendo proporcionalmente,

$$\hat{p}(\mathbf{x}, C_c) = \frac{k_c}{KV_R}$$

es razonable tomar

$$\hat{\Pr}(C_c | \mathbf{x}) = \frac{\hat{p}(\mathbf{x}, C_c)}{\hat{p}(\mathbf{x})} = \frac{k_c}{k}$$

C/DTC/UCIIM

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

ntanas de Parzen

Procediendo análogamente

$$\hat{p}(\mathbf{x}) = \frac{1}{Kh^N} \sum_{k'=1}^K \phi\left(\frac{\mathbf{x} - \mathbf{x}^{(k')}}{h}\right)$$

$$\hat{p}(\mathbf{x}, C_c) = \frac{1}{Kh^N} \sum_{k'=1}^{K_c} \phi\left(\frac{\mathbf{x} - \mathbf{x}_c^{(k')}}{h}\right)$$

\mathbf{x}_c indica pertenencia a C_c , y K_c es el número de muestras de la clase c ; a de ello

$$\hat{\Pr}(C_c | \mathbf{x}) = \frac{\sum_{k'=1}^{K_c} \phi\left(\frac{\mathbf{x} - \mathbf{x}_c^{(k')}}{h}\right)}{\sum_{k'=1}^K \phi\left(\frac{\mathbf{x} - \mathbf{x}^{(k')}}{h}\right)}$$

C/DTC/UCIIM

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

imadores semiparamétricos (mezclas)

Resulta paralelo al caso de las ventanas; así, para mezclas de gaussianas,

$$\hat{p}(\mathbf{x}, C_c) = \sum_{j=1}^{J_c} \rho_c^{(j)} \frac{1}{(2\pi)^{N/2} |\mathbf{V}_c^{(j)}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m}_c^{(j)})^T [\mathbf{V}_c^j]^{-1} (\mathbf{x} - \mathbf{m}_c^{(j)})\right\}$$

o

$$C_c | \mathbf{x} = \frac{\sum_{j=1}^{J_c} \rho_c^{(j)} \frac{1}{(2\pi)^{N/2} |\mathbf{V}_c^{(j)}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m}_c^{(j)})^T [\mathbf{V}_c^j]^{-1} (\mathbf{x} - \mathbf{m}_c^{(j)})\right\}}{\sum_{c=1}^C \sum_{j=1}^{J_c} \rho_c^{(j)} \frac{1}{(2\pi)^{N/2} |\mathbf{V}_c^{(j)}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m}_c^{(j)})^T [\mathbf{V}_c^j]^{-1} (\mathbf{x} - \mathbf{m}_c^{(j)})\right\}}$$

C/DTC/UCHIM



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

Mo de Ampliación

ón vía estimación NP de las ddp

Una vez obtenidas las $\hat{Pr}(C_c | \mathbf{x})$, puede aplicarse la Teoría Bayesiana para decidir.

Establezca los modos de hacerlo para situaciones MAP a partir de los casos anteriores.

¿Cuál es la principal característica de estos métodos?

Al tratarse de una situación MAP, el proceso a seguir será

$$c^* = \arg \left[\max_c \{ Pr(C_c | \mathbf{x}) \} \right]$$

do que se tiene:

k-NN:

$$c^* = \arg \left[\max_c \{ k_c \} \right]$$

C/DTC/UCIIM

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

Parzen:

$$c^* = \arg \left[\max_c \left\{ \sum_{k'=1}^K \phi \left(\frac{\mathbf{x} - \mathbf{x}_c^{(k')}}{h} \right) \right\} \right]$$

Mezclas gaussianas:

$$\arg \left[\max_c \left\{ \sum_{j=1}^{J_c} \rho_c^{(j)} \frac{1}{(2\pi)^{N/2} |\mathbf{V}_c^{(j)}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}_c^{(j)})^T [\mathbf{V}_c^{(j)}]^{-1} (\mathbf{x} - \mathbf{m}_c^{(j)}) \right\} \right\} \right]$$

La característica de estos métodos es que ninguno se diseña orientado a la decisión, sino a partir de la estimación de las ddp: lo que es especialmente relevante en el caso semiparamétrico, ya que los $\rho_c^{(j)}$, $\mathbf{m}_c^{(j)}$ y $\mathbf{V}_c^{(j)}$ se establecen con un objetivo distinto de la propia clasificación (ML, típicamente). No obstante, la decisión k -NN, por su sencillez y fácil interpretación, es un método muy utilizado.

C/DTC/UCIIM

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

Estimación de Nadaraya-Watson

Si se aproxima la ddp de (\mathbf{s}, \mathbf{x}) (dimensiones N_1, N_2 , respectivamente) mediante ventanas de Parzen gaussianas circulares (factorizables) a partir de K (de) muestras, se tendrá:

$$p(\mathbf{s}, \mathbf{x}) = \frac{1}{K} \sum_{k'=1}^K \frac{1}{(2\pi v)^{(N_1+N_2)/2}} \exp\left(-\frac{1}{2v} \|\mathbf{s} - \mathbf{s}^{(k')}\|_2^2\right) \exp\left(-\frac{1}{2v} \|\mathbf{x} - \mathbf{x}^{(k')}\|_2^2\right)$$

partir de ello, se aplica la estimación MS, que implica elegir

$$E\{\mathbf{s} | \mathbf{x}\} = \int \mathbf{s} p(\mathbf{s} | \mathbf{x}) d\mathbf{s} = \int \frac{\mathbf{s} p(\mathbf{s}, \mathbf{x})}{p(\mathbf{x})} d\mathbf{s} = \frac{\int \mathbf{s} p(\mathbf{s}, \mathbf{x}) d\mathbf{s}}{\int p(\mathbf{s}, \mathbf{x}) d\mathbf{s}}$$

UC/DTIC/UCIIM

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

se tiene

$$\hat{\mathbf{s}}_p(\mathbf{x}) = \frac{\int \mathbf{s} \hat{p}(\mathbf{s}, \mathbf{x}) d\mathbf{s}}{\int \hat{p}(\mathbf{s}, \mathbf{x}) d\mathbf{s}} = \frac{\sum_{k'=1}^K \mathbf{s}^{(k')} \exp\left(-\frac{1}{2\mathbf{v}} \|\mathbf{x} - \mathbf{x}^{(k')}\|_2^2\right)}{\sum_{k'=1}^K \exp\left(-\frac{1}{2\mathbf{v}} \|\mathbf{x} - \mathbf{x}^{(k')}\|_2^2\right)}$$

Análogamente se puede proceder con mezclas.

El inconveniente es el mismo que el considerado en decisión: no hay solución del planteamiento a optimizar la estimación de \mathbf{s} .



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

C/DTC/UCIIM