Estimación de densidad de probabilidad mediante ventanas de Parzen

Pedro J. García Laencina*, José Luis Sancho Gómez.
Departamento de Tecnologías de la Información y las Comunicaciones
Universidad Politécnica de Cartagena. Plaza del Hospital 1, 30202 Cartagena
Teléfono: (+34) 968326542. E-mail: pedroj.garcia@upct.es

Resumen. *Este trabajo presenta la estimación de funciones de densidad de probabilidad mediante ventanas de Parzen, que constituye una de las técnicas no-paramétricas más extendidas en este campo. Se analizan experimentalmente sus capacidades en un problema de procesado de imagen.*

1. Introducción

La estimación de funciones de densidad de probabilidad (fdp) es necesaria en multitud de escenarios y aplicaciones reales, como son el reconocimiento de patrones, el registro de imagen y la segmentación de imágenes. En la literatura destaca la técnica no-paramétrica conocida como *Ventanas de Parzen* [1], [2], [3]. Este artículo presenta y analiza esta extendida técnica.

2. Estimación no-paramétrica de densidades

Supóngase que se dispone de un conjunto de N patrones d-dimensionales definido por $\mathcal{X} = (x_{in})$, que es una matriz de datos rectangular de dimensiones $d \times N$, siendo x_{in} el valor de la característica i-ésima para el patrón $\mathbf{x}_n$. El objetivo es modelar la fdp que generó los datos, $p(\mathbf{x})$, sin asumir previamente ninguna forma determinada para la fdp. Estas técnicas no-paramétricas se fundamentan en que la probabilidad de que un nuevo vector $\mathbf{x}$, obtenido a partir de una fdp desconocida $p(\mathbf{x})$, caiga dentro de alguna región $\mathcal{R}$ del espacio de entrada viene, por definición, dada por

$$P = \int_{\mathcal{R}} p(\mathbf{x}')\, d\mathbf{x}'. \tag{1}$$

Si se dispone de N muestras obtenidas independientemente a partir de $p(\mathbf{x})$, se puede obtener una buena estimación de la probabilidad P a partir de la fracción media de muestras que caen en $\mathcal{R}$, de forma que

$$P \approx K/N. \tag{2}$$

Además, si se asume que $p(\mathbf{x})$ es continua y que no varía apreciablemente sobre la región $\mathcal{R}$, entonces es posible aproximar (1) por

$$P = \int_{\mathcal{R}} p(\mathbf{x}')\, d\mathbf{x}' \approx p(\mathbf{x})V, \tag{3}$$

donde V es el volumen de $\mathcal{R}$, y $\mathbf{x}$ es un patrón incluido en $\mathcal{R}$. De (2) y (3), se obtiene

$$p(\mathbf{x}) \approx \frac{K}{NV}. \tag{4}$$

Atendiendo a la estimación de densidad dada por (4), se pueden adoptar dos métodos básicos. El primero consiste en elegir un valor fijo para K y determinar V a partir de los datos. Alternativamente, se puede fijar el volumen V y determinar K a partir de los datos. Esto nos lleva a las técnicas de estimación de densidad tipo '*Kernel*', que se describen a continuación.

3. Ventanas de Parzen

Supóngase una región $\mathcal{R}$ definida por un hipercubo con lados de longitud h centrados en el punto $\mathbf{x}$. Entonces su volumen viene dado por

$$V = h^d. \tag{5}$$

Podemos encontrar una expresión para K, el número de muestras que caen en esta región, definiendo una función *Kernel* o núcleo $\phi(\mathbf{u})$, también conocida como *Ventana básica de Parzen* [3] dada por

$$\phi(\mathbf{u}) = \begin{cases} 1 & |u_i| < 1/2 \\ 0 & \text{otro caso.} \end{cases} \tag{6}$$

De este modo, $\phi(\mathbf{u})$ se corresponde con un cubo unidad centrado en el origen. Por tanto, para cada $\mathbf{x}_n$, la cantidad $\phi\left((\mathbf{x} - \mathbf{x}_n)/h\right)$ es igual a la unidad si $\mathbf{x}_n$ cae dentro del hipercubo de lado h centrado en $\mathbf{x}$, y es cero si no es así. En la literatura, h se conoce como *parámetro de suavizado* o *ancho de kernel*. El número total de muestras que caen dentro del hipercubo es simplemente

$$K = \sum_{n=1}^{N} \phi\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right). \tag{7}$$

Si se sustituye (5) y (7) en (4), se obtiene la siguiente estimación para la densidad en $\mathbf{x}$:

$$\hat{p}\,(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{h^d} \phi\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right) \tag{8}$$

donde $\hat{p}(\mathbf{x})$ denota la densidad estimada mediante ventanas de Parzen [3]. Esta estimación de fdp puede verse como la superposición de N cubos de lado h, con un hipercubo centrado en cada una de las muestras. Este método es similar a la estimación basada en el histograma, excepto porque en vez de intervalos se tienen hipercubos cuya posición está determinada por los datos. Sin embargo, sigue presente el problema de las discontinuidades en la estimación de la fdp [1], [2]. Para solucionarlo, una opción muy utilizada es emplear un núcleo gaussiano:

$$G(\mathbf{x}, \mathbf{x}_n, h) = \frac{1}{(2\pi h^2)^{d/2}} e^{-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{2h^2}} \tag{9}$$

donde h representa la desviación estandar en cada dimensión de entrada. De esta forma, la estimación de la fdp mediante Parzen queda

$$\hat{p}(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} G(\mathbf{x}, \mathbf{x}_n, h), \tag{10}$$

En general, si las funciones de núcleo satisfacen

$$\phi(\mathbf{u}) \geq 0 \tag{11}$$

y

$$\int \phi(\mathbf{u})\, d\mathbf{u} = 1 \tag{12}$$

entonces la estimacion de (8) satisface que $\hat{p}(\mathbf{x}) > 0$ y $\int \hat{p}(\mathbf{x})\ d\mathbf{x} = 1$.

4. Cálculo de *h*

Establecer un valor para h no es una tarea sencilla, ya que su valor óptimo h_{opt} (es decir, el valor que minimiza la diferencia entre $\hat{p}(\mathbf{x})$ y $p(\mathbf{x})$) depende en gran medida de la naturaleza de los datos de entrada y el número de patrones.

4.1. Criterio de Silverman

Silverman propuso en [4] la siguiente regla general de carácter práctico para calcular el valor de h:

$$h_{SIL} = \frac{0{,}9A}{N^{1/5}} \tag{13}$$

siendo $A = \text{mín}\left\{s, \frac{r}{1{,}34}\right\}$, donde s es la desviación estándar y r es el rango intercuartil, medidos en el conjunto de datos.

4.2. Validación cruzada de orden uno

Un procedimiento muy extendido es obtener el valor óptimo de h mediante validación cruzada de orden uno o *'Leave-One-Out'* (LOO). La estimación LOO de ventanas de Parzen viene dada por

$$\hat{p}_{-n}(\mathbf{x}_n) = \frac{1}{N-1} \sum_{\substack{m=1 \\ m \neq n}}^{N} \frac{1}{(2\pi h^2)^{d/2}} e^{-\frac{\|\mathbf{x}-\mathbf{x}_m\|^2}{2h^2}} \tag{14}$$

que representa la estimación de la fdp a partir de $N-1$ patrones de entrenamiento excluyendo $\mathbf{x}_n$ [5].

4.2.1. Maximización de la verosimilitud: Según el criterio ML (*'Maximum Likelihood'*), se debe escoger aquel valor de h que maximice la verosimilitud del conjunto de datos:

$$\mathcal{L}(h) \equiv \prod_{n=1}^{N} p(\mathbf{x}_n|h), \tag{15}$$

que es equivalente a minimizar

$$E_{ML}(h) = -\ln \mathcal{L}(h) = -\sum_{n=1}^{N} \ln p(\mathbf{x}_n|h). \tag{16}$$

Sustituyendo el estimador LOO dado por (14) en (16), se obtiene el siguiente criterio:

$$E_{ML}^{LOO}(h) = -\sum_{n=1}^{N} \ln \hat{p}_{-n}(\mathbf{x}_n|h). \tag{17}$$

Así, si se realiza un barrido de h dentro de un cierto rango, su valor óptimo puede obtenerse mediante

$$h_{ML} = \arg\min_{h} E_{ML}^{LOO}(h). \tag{18}$$

5. Ejemplo: Una imagen digital

Una vez se ha descrito el método de ventanas de Parzen, se utiliza una imagen digital para presentar la capacidad de esta técnica. En concreto, la imagen utilizada es una fotografía reciente del Cuartel de Antigones (Cartagena, España), sede actual de la ETSIT de la Universidad Politécnica de Cartagena —ver Figura 1(a)—. Esta imagen tiene 256 niveles de gris y un tamaño de 256x256 píxeles. Se ha representado el histograma de dicha imagen en la Figura 1(b). Hay que destacar que para imágenes digitales se dispone de variables discretas (con 256 valores, en este caso) y, por tanto, no se puede hablar de una función de densidad (asociadas a variables continuas), sino de una función de probabilidad que viene dada por el histograma. Sin embargo, en procesamiento de imagen suele ser útil considerar el cálculo de la fdp para poder disponer de una función continua en lugar del histograma. Inicialmente, se evalua el efecto del parámetro h y, a continuación, se calcula h_{opt} mediante el criterio de Silverman [4].

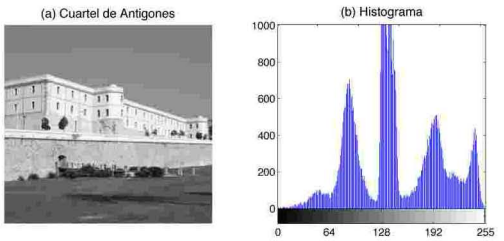

Fig. 1. En (a) se muestra la imagen analizada; mientras que (b) muestra su histograma.

En general, cuando se dispone de un número suficiente de muestras, se verifica que la estimación de Parzen con $h \to 0$ consigue que $\hat{p}(\mathbf{x}) \approx p(\mathbf{x})$. Por tanto, en este caso, la estimación obtenida debe tender al histograma cuando $h \to 0$.

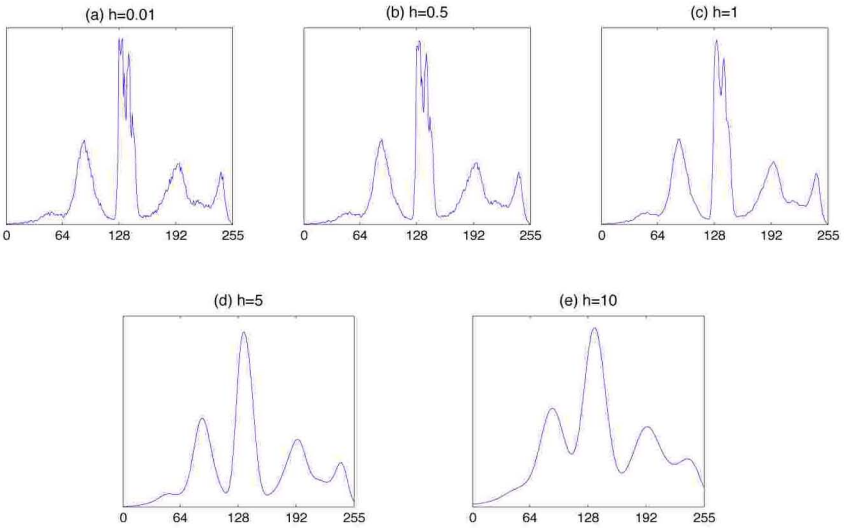

Fig. 2. Estimación de fdp mediante ventanas de Parzen para la imagen considerada utilizando distintos valores de h.

La Figura 2 muestra la estimación obtenida utilizando los siguientes valores de h: $1e-1$, $5e-1$, 1, 5 y 10. Se cumple que para valores bajos de h se consigue una estimación muy ruidosa y con discontinuidades (próxima al histograma), mientras que es más suave conforme aumenta el tamaño de la ventana. En la práctica, es necesario llegar a un compromiso entre ambas situaciones.

Seguidamente, la Figura 3 muestra la evolución del error E_{ML}^{LOO} con respecto de h. En particular, se ha realizado un barrido del valor de h entre 0,1 y 10, con incrementos de 0,1. Como es natural, y dada la naturaleza discreta de la imagen considerada y el sufiente número de casos que se dispone, se puede comprobar que la mejor aproximación a la solución teórica (histograma) se consigue con valores muy bajos de h y se verifica que el error incrementa conforme aumenta h.

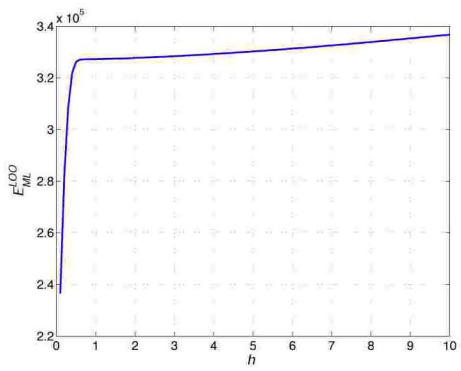

Fig. 3. Criterio ML: Evolución del error con respecto de h.

A continuación, se utiliza el criterio de Silverman para calcular el valor de h, obteniendo $h_{SIL} = 5{,}09$. En muchos escenarios y aplicaciones prácticas, este criterio constituye un procedimiento sencillo para obtener una estimación suficiente y adecuada para la función de densidad de probabilidad.

6. Conclusiones

El cálculo de fdps se realiza en múltiples aplicaciones de procesado de señal e imágenes. Dentro de las distintas técnicas existentes, una de las más utilizadas es el método de ventanas de Parzen. En esta técnica no-paramétrica la estimación de la fdp viene definido por el conjunto de datos y un único parámetro h —el tamaño de la ventana—. Se ha presentado un procedimiento para obtener un valor óptimo para h y se ha mostrado su utilidad en la estimación de la fdp de una imagen digital.

Referencias

[1] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, USA: Oxford University Press (1995).
[2] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.
[3] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
[4] B. W. Silverman. *Density Estimation*. Chapman & Hall, 1986.
[5] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer, Berlin, 2004.