Ventanas de Parzen

Ventanas de Parzen

Tags: Reconocimiento de Patrones, Estimación no paramétrica de funciones de densidad de probabilidad

---

Suponga que se tiene un conjunto de datos de N patrones d-dimensionales definido por $x = (x_{in})$, que es una matriz de tamaño $d \times N$, siendo x_{in} el valor de la característica i-esima para el patron x_n.

El objetivo es modelar la PDF que generó los datos $p(x)$, sin asumir nada previamente respecto a la forma de la PDF.

Estas técnicas no parametricas se fundamentan en que la probabilidad de un nuevo vector x, obtenido a partir de una PDF desconocida $p(x)$, caiga dentro de alguna region R del espacio de entrada, dada por:

$$P = \int_R p(x')dx' \qquad (2.4.1)$$

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

$$P \approx k/N \qquad (2.4.2)$$

Además, se asume que $p(x)$ es continua y no varía apreciablemente sobre la región R, entonces es posible aproximar $(2.4.1)$ por:

$$P = \int_R p(x')dx' \approx p(x)V \qquad (2.4.3)$$

donde V es el hipervolumen de R y x es un patrón incluido en R.

De $(2.4.2)$ y $(2.4.3)$, se obtiene

$$k/N = p(x)V$$

$$p(x) = \frac{k}{NV} \quad (2.4.4)$$

Suponga una región R definida por un hipercubo con lados de longitud h, centrados en el punto x. Entonces su volumen viene dado por:

$$V = h^d \qquad (2.4.5)$$

podemos encontrar una expresión para k, el número de muestras que caen en esta región, definiendo una función kernel ($\phi(u)$), también conocida como ventana básica de Parzen la cual está dada por:

$$\phi(\mathrm{u}) = \begin{cases} 1 & |\mathrm{u_j}| \leq \frac{1}{2} \qquad \mathrm{j} = 1, \ldots, \mathrm{d} \\ 0 & \text{en otro caso} \end{cases} \qquad (2.4.6)$$

De este modo, $\phi(u)$ se corresponde con un cubo unidad centrado en el origen. Por tanto, para cada x_n, la cantidad $\phi(x - x_n)/h$ es igual a la unidad si x_n cae dentro del hipercubo de lado h centrado en x y es 0 si no es así.

En la literatura, h se conoce como un parámetro de suavizado, o el ancho del kernel. El número total de muestras que caen dentro del hipercubo es

$$K = \sum_{n=1}^{N} \phi(\frac{x - x_n}{h}) \quad (2.4.7)$$

Si se sustituye $(2.4.5)$ y $(2.4.7)$ en $(2.4.4)$ se obtiene la siguiente estimación de la densidad en x

$$p(x) = \frac{k}{NV} = \frac{k}{Nh^d}$$

$$p(x) = \frac{1}{N} \sum_{n=1}^{N} \phi(\frac{x - x_n}{h}) \frac{1}{h^d} \qquad (2.4.8)$$

donde $\hat{p}(x)$ es la PDF estimada.

Esta estimación de $\hat{p}(x)$ puede verse como la superposición de N cubos de lado h, con un hipercubo centrado en cada una de las muestras.

Por el problema de las discontinuidades podemos usar un kernel gaussiano.

$$G(x, x_n, h) = \frac{1}{(2\pi h^2)^{d/2}} e^{-\frac{||x - x_n||^2}{2h^2}} \qquad (2.4.9)$$

donde h es la desviación estándar en cada dimensión de entrada.

$$G(x, x_n, h) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-1/2(x - x_n)^T \Sigma^{-1} (x - x_n)} \qquad (2.4.10)$$

entonces:

$$\hat{p}(x) = \frac{1}{N}\sum_{n=1}^{N} G(x, x_n, \Sigma) \qquad (2.4.11)$$

$$\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(v_{x_j}-\mu_{x_i})^2}{2\sigma_1^2}\right)$$

References

Estimación de funciones de densidad de probabilidad por k-vecinos más cercanos