

Métodos No Lineales

Alvaro J. Riascos Villegas
Universidad de los Andes y Quantil

Junio de 2019

Contenido

- 1 Métodos de Kernels: Aprendizaje no supervisado
- 2 Clasificación

Estimación de densidades

- Supongamos que x_1, \dots, x_N es una muestra de datos tomada con una distribución con densidad $f_X(x)$.
- Un primer estimador local es:

$$f_X(x_0) = \frac{\text{num}\{N(x_0)\}}{N\lambda} \quad (1)$$

donde $N(x_0)$ es una vecindad de tamaño λ .

- Una versión suavizada es:

$$f_X(x_0) = \frac{1}{N\lambda} \sum_{i=1}^N K_\lambda(x_0, x_i) \quad (2)$$

donde por ejemplo $K_\lambda(x_0, x_i) = \psi_\lambda(x - x_0)$ y ψ_λ es la densidad Gaussiana $(0, \lambda^2)$.

Estimación de densidades

- Supongamos que x_1, \dots, x_N es una muestra de datos tomada con una distribución con densidad $f_X(x)$.
- Un primer estimador local es:

$$f_X(x_0) = \frac{\text{num}\{N(x_0)\}}{N\lambda} \quad (1)$$

donde $N(x_0)$ es una vecindad de tamaño λ .

- Una versión suavizada es:

$$f_X(x_0) = \frac{1}{N\lambda} \sum_{i=1}^N K_\lambda(x_0, x_i) \quad (2)$$

donde por ejemplo $K_\lambda(x_0, x_i) = \psi_\lambda(x - x_0)$ y ψ_λ es la densidad Gaussiana $(0, \lambda^2)$.

Estimación de densidades

- Supongamos que x_1, \dots, x_N es una muestra de datos tomada con una distribución con densidad $f_X(x)$.
- Un primer estimador local es:

$$f_X(x_0) = \frac{\text{num}\{N(x_0)\}}{N\lambda} \quad (1)$$

donde $N(x_0)$ es una vecindad de tamaño λ .

- Una versión suavizada es:

$$f_X(x_0) = \frac{1}{N\lambda} \sum_{i=1}^N K_\lambda(x_0, x_i) \quad (2)$$

donde por ejemplo $K_\lambda(x_0, x_i) = \psi_\lambda(x - x_0)$ y ψ_λ es la densidad Gaussiana $(0, \lambda^2)$.

Estimación de densidades

- En forma reducida:

$$f_X(x) = \frac{1}{N\lambda} \sum_{i=1}^N \psi_\lambda(x - x_i) \quad (3)$$

- En p dimensiones:

$$f_X(x) = \frac{1}{N\lambda(2\lambda^2\pi)^{\frac{p}{2}}} \sum_{i=1}^N \exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\lambda}\right)^2\right) \quad (4)$$

Estimación de densidades

- En forma reducida:

$$f_X(x) = \frac{1}{N\lambda} \sum_{i=1}^N \psi_\lambda(x - x_i) \quad (3)$$

- En p dimensiones:

$$f_X(x) = \frac{1}{N\lambda(2\lambda^2\pi)^{\frac{p}{2}}} \sum_{i=1}^N \exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\lambda}\right)^2\right) \quad (4)$$

Estimación de densidades

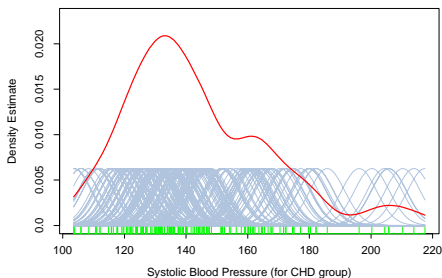


FIGURE 6.13. A kernel density estimate for systolic blood pressure (for the CHD group). The density estimate at each point is the average contribution from each of the kernels at that point. We have scaled the kernels down by a factor of 10 to make the graph readable.

Contenido

- 1 Métodos de Kernels: Aprendizaje no supervisado
- 2 Clasificación

Clasificación

- Usando las estimaciones de Kernel de las densidades por clase f_j , el clasificador óptimo se puede escribir (usando la regla de Bayes):

$$P(G = j | X = x_0) = \frac{\pi_j f_j(x_0)}{\sum_{k=1}^J \pi_k f_k(x_0)} \quad (5)$$

donde π_j son las frecuencias relativas de cada clase.

Estimación de densidades

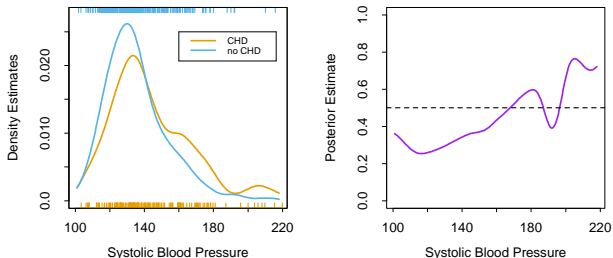


FIGURE 6.14. *The left panel shows the two separate density estimates for systolic blood pressure in the CHD versus no-CHD groups, using a Gaussian kernel density estimate in each. The right panel shows the estimated posterior probabilities for CHD, using (6.25).*

Clasificación: Bayes Naive

- Cuando el espacio de características es muy grande (p grande) la estimación por kernels tiene mucha varianza.
- El clasificador de Bayer asume independencia de las variables y en ese caso el clasificador de Bayes se reduce a:

$$P(G = j | X = x_0) = \frac{\pi_j f_j(x_0)}{\sum_{k=1}^J \pi_k f_k(x_0)} \quad (6)$$

donde:

$$f_j(x) = \prod_{l=1}^p f_{j,l}(x_l) \quad (7)$$

- Cada $f_{j,l}$ es un kernel unidimensional. La complejidad del problema se reduce enormemente.
- Si alguna variable es discreta, permite combinar fácilmente variables continuas y categóricas

Clasificación: Bayes Naive

- Cuando el espacio de características es muy grande (p grande) la estimación por kernels tiene mucha varianza.
- El clasificador de Bayer asume independencia de las variables y en ese caso el clasificador de Bayes se reduce a:

$$P(G = j | X = x_0) = \frac{\pi_j f_j(x_0)}{\sum_{k=1}^J \pi_k f_k(x_0)} \quad (6)$$

donde:

$$f_j(x) = \prod_{l=1}^p f_{j,l}(x_l) \quad (7)$$

- Cada $f_{j,l}$ es un kernel unidimensional. La complejidad del problema se reduce enormemente.
- Si alguna variable es discreta, permite combinar fácilmente variables continuas y categóricas

Clasificación: Bayes Naive

- Cuando el espacio de características es muy grande (p grande) la estimación por kernels tiene mucha varianza.
- El clasificador de Bayer asume independencia de las variables y en ese caso el clasificador de Bayes se reduce a:

$$P(G = j | X = x_0) = \frac{\pi_j f_j(x_0)}{\sum_{k=1}^J \pi_k f_k(x_0)} \quad (6)$$

donde:

$$f_j(x) = \prod_{l=1}^p f_{j,l}(x_l) \quad (7)$$

- Cada $f_{j,l}$ es un kernel unidimensional. La complejidad del problema se reduce enormemente.
- Si alguna variable es discreta, permite combinar fácilmente variables continuas y categóricas

Clasificación: Bayes Naive

- Cuando el espacio de características es muy grande (p grande) la estimación por kernels tiene mucha varianza.
- El clasificador de Bayer asume independencia de las variables y en ese caso el clasificador de Bayes se reduce a:

$$P(G = j | X = x_0) = \frac{\pi_j f_j(x_0)}{\sum_{k=1}^J \pi_k f_k(x_0)} \quad (6)$$

donde:

$$f_j(x) = \prod_{l=1}^p f_{j,l}(x_l) \quad (7)$$

- Cada $f_{j,l}$ es un kernel unidimensional. La complejidad del problema se reduce enormemente.
- Si alguna variable es discreta, permite combinar fácilmente variables continuas y categóricas.