



Profesores: Carlos Hurtado, Pablo Barceló
Auxiliar: Gonzalo Ríos
Fecha: 13 de Agosto

Auxiliar 2: Introducción a la Clasificación

1 Materia

1. Se define el problema de clasificación como sigue:

- $X = (X_1, \dots, X_n)$ variables de entrada
- C variable de la clase
- $\text{dom}(C) = \{C_1, \dots, C_k\}$, clases
- Encontrar una función $f : \text{dom}(X) \rightarrow \text{dom}(C)$, llamada clasificador

2. Existen varios enfoques del problema, entre ellos:

- Probabilístico: $p : \text{dom}(X) \times \text{dom}(C) \rightarrow [0, 1]$, probabilidad conjunta
 - (a) Bayes-Naive
 - (b) Redes Bayesianas
 - (c) Clasificadores paramétricos
- Regresión: $f(x) = c_k$ si $f_k(x) > 0$. Los f_k minimizan un cierto criterio (mínimos cuadrados)
 - (a) Lineal
 - (b) Logístico
 - (c) Redes Neuronales
- Discriminante: inducimos directamente los límites de las clases
 - (a) Árboles de decisión
 - (b) Reglas de decisión
 - (c) Discriminantes lineales

3. Propiedades probabilísticas

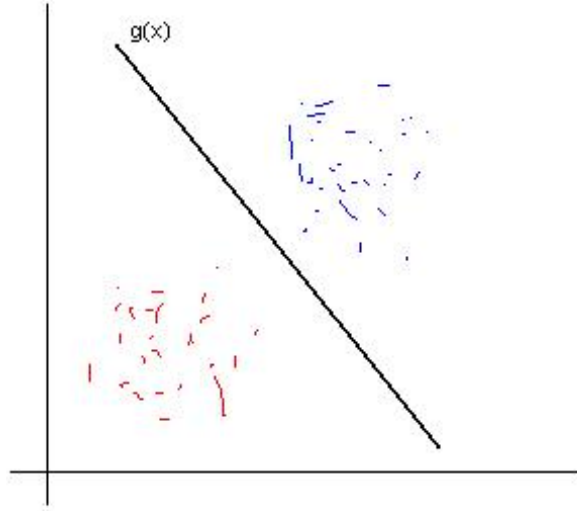
- (a) Probabilidad condicional: $\Pr(A|B) = \frac{\Pr(A, B)}{\Pr(B)}$
- (b) Regla de marginación: $\Pr(A) = \sum \Pr(A, B_i)$, B_i disjuntos, $\bigcup B_i = \Omega$
- (c) Teorema de Bayes: $\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)}$
- (d) Independencia: X, Y v.a indep ssi $\forall x, y, \Pr(X = x, Y = y) = \Pr(X = x) \times \Pr(Y = y)$

2 Problemas

2.1 Inducción de un discriminante lineal

Sean $(\vec{x}_i, c_i)_{i=1}^k$ datos, donde $\text{dom}(C) = \{-1, 1\}$ y $\vec{x}_i = (x_{1i}, \dots, x_{ni})$. Plantee el problema de minimización que induzca un plano que separe las clases, de la mejor forma posible. Suponga que las clases son linealmente separables.

Queremos encontrar $g(x_1, \dots, x_n) = a_0 + a_1x_1 + \dots + a_nx_n$ tal que $f(x_{1i}, \dots, x_{ni}) = \begin{cases} 1 & \text{si } g(x_{1i}, \dots, x_{ni}) > 0 \\ -1 & \text{si } g(x_{1i}, \dots, x_{ni}) \leq 0 \end{cases}$.



La distancia entre un punto $(x_{1i}, \dots, x_{ni})$ y el plano $g(x_1, \dots, x_n)$ es

$$d(\vec{x}_i, g(\vec{x})) = \frac{|g(x_{1i}, \dots, x_{ni})|}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2}}$$

ya que la proyección de $(x_{1i}, \dots, x_{ni})$ en el plano $g(x_1, \dots, x_n)$ es tal que

$$\min \frac{1}{2} |x_i - \bar{x}|^2, \text{ s.a. } g(\bar{x}) = 0$$

$$\begin{aligned} \implies \min \frac{1}{2} \sum_{k=1}^n (x_{ki} - \bar{x}_k)^2 - \lambda a_k \bar{x}_k &\implies \text{derivando, } -(x_{ki} - \bar{x}_k) - \lambda a_k = 0 \implies \bar{x}_k = \lambda a_k + x_{ki} \\ \implies a_0 + a_1 \left(\frac{\lambda}{2} a_1 + x_{1i} \right) + \dots + a_n (\lambda a_n + x_{ni}) &= 0 \implies a_0 + a_1 x_{1i} + \dots + a_n x_{ni} + \lambda (a_1^2 + \dots + a_n^2) = 0 \\ \implies \lambda = -\frac{a_0 + a_1 x_{1i} + \dots + a_n x_{ni}}{(a_1^2 + \dots + a_n^2)} & \\ |x_i - \bar{x}| = |\lambda| \sqrt{(a_1^2 + \dots + a_n^2)} &= \frac{|g(x_{1i}, \dots, x_{ni})|}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2}} \end{aligned}$$

Luego, $g(x)$ debe cumplir que

$$\begin{aligned} \min \frac{1}{2} \sum_{i=1}^k \frac{(a_0 + a_1 x_{1i} + \dots + a_n x_{ni})^2}{a_1^2 + a_2^2 + \dots + a_n^2} \\ \text{s.a. } g(\vec{x}_i) \times c_i > 0 \end{aligned}$$

2.2 Clasificador Bayesiano Discreto

Pedro, Juan y Diego son tres amigos coleccionistas de estampillas. Todas las semanas van juntos a la tienda de filatelia para comprar nuevas estampillas. Ellos tienen diversos gustos, y la colección de ellos se resume en:

	Animales	Países	Historia	Total
Pedro	10	50	100	160
Juan	50	10	80	140
Diego	100	30	20	150
Total	160	90	200	450

Esta semana, la tienda de filatelia tenía una estampilla especial de un Caballo Histórico, ¿Quién de los tres amigos comprará la estampilla? Asuma independencia en los gustos de cada amigo

En este problema, $C = \{Pedro, Juan, Diego\}$, $X = \{A, P, H\}$, $Dom(X) = \{0, 1\} \times \{0, 1\} \times \{0, 1\}$ y $x = (1, 0, 1)$.

$$\Pr(C|A, P, H) = \frac{\Pr(A, P, H|C) \times \Pr(C)}{\Pr(A, P, H)} = \frac{\Pr(A|C) \times \Pr(P|C) \times \Pr(H|C) \times \Pr(C)}{\Pr(A, P, H)}$$

Calculemos el valor de las funciones discriminantes:

1. Pedro: $g_P(1, 0, 1) = \Pr(A = 1|Pedro) \times \Pr(P = 0|Pedro) \times \Pr(H = 1|Pedro) \times \Pr(Pedro)$
 $= \frac{10}{160} \times \frac{110}{160} \times \frac{100}{160} \times \frac{160}{450} = \frac{11}{1152} = 9.5486 \times 10^{-3}$
2. Juan: $g_J(1, 0, 1) = \Pr(A = 1|Juan) \times \Pr(P = 0|Juan) \times \Pr(H = 1|Juan) \times \Pr(Juan)$
 $= \frac{50}{140} \times \frac{130}{140} \times \frac{80}{140} \times \frac{140}{450} = \frac{26}{441} = 5.8957 \times 10^{-2}$
3. Diego: $g_D(1, 0, 1) = \Pr(A = 1|Diego) \times \Pr(P = 0|Diego) \times \Pr(H = 1|Diego) \times \Pr(Diego)$
 $= \frac{100}{150} \times \frac{120}{150} \times \frac{20}{150} \times \frac{150}{450} = \frac{16}{675} = 2.3704 \times 10^{-2}$

Juan comprará la estampilla de esta semana.

2.3 Clasificador Bayesiano Continuo

Se desea pronosticar si un día va a llover o no. Las variables que se disponen son la temperatura y la humedad. Los registros de los días anteriores son los siguientes:

T	H	$Lluvia$
15	95	Si
25	70	No
10	80	No
10	90	Si

Si las mediciones de hoy son (14, 85), determine si va a llover o no. Use estimación normal y asuma independencia para T y H .

Debemos calcular $\Pr(Lluvia|T = 14, H = 85)$ y $\Pr(No Lluvia|T = 14, H = 85)$. Para ello, estimemos las funciones de densidad condicionales:

1. Temperatura:

$$\bullet \text{ Lluvia: } \bar{T} = \frac{15+10}{2} = 12.5, S_T^2 = \frac{(15-12.5)^2 + (10-12.5)^2}{2-1} = 2.5^2 + 2.5^2 = 12.5$$

$$T|Lluvia \sim N(12.5, 12.5)$$

$$\bullet \text{ No Lluvia: } \bar{T} = \frac{25+10}{2} = 17.5, S_T^2 = \frac{(25-17.5)^2 + (10-17.5)^2}{2-1} = 112.5$$

$$T|No Lluvia \sim N(17.5, 112.5)$$

2. Humedad:

$$\bullet \text{ Lluvia: } \bar{H} = \frac{95+90}{2} = 92.5, S_H^2 = \frac{(95-92.5)^2 + (90-92.5)^2}{2-1} = 12.5$$

$$H|Lluvia \sim N(92.5, 12.5)$$

$$\bullet \text{ No Lluvia: } \bar{H} = \frac{70+80}{2} = 75, S_H^2 = \frac{(70-75)^2 + (80-75)^2}{2-1} = 50.0$$

$$H|No Lluvia \sim N(75, 50)$$

Ahora calculemos los valores discriminantes:

1. Si *Lluvia*

$$\begin{aligned} g_{Si}(T=14, H=85) &= f_T(T=14|Lluvia) \times f_H(H=85|Lluvia) \times \Pr(Lluvia) \\ &= \frac{1}{\sqrt{2\pi \times 12.5}} \exp\left(-\frac{(14-12.5)^2}{2 \times 12.5}\right) \times \frac{1}{\sqrt{2\pi \times 92.5}} \exp\left(-\frac{(85-92.5)^2}{2 \times 92.5}\right) \times 0.5 = 9.7934 \times 10^{-16} \end{aligned}$$

2. No *Lluvia*

$$\begin{aligned} g_{No}(T=14, H=85) &= f_T(T=14|No Lluvia) \times f_H(H=85|No Lluvia) \times \Pr(No Lluvia) \\ &= \frac{1}{\sqrt{2\pi \times 112.5}} \exp\left(-\frac{(14-17.5)^2}{2 \times 112.5}\right) \times \frac{1}{\sqrt{2\pi \times 50}} \exp\left(-\frac{(85-75)^2}{2 \times 50}\right) \times 0.5 = 3.6965 \times 10^{-4} \end{aligned}$$

Luego, hoy no va a llover.