El Clasificador Bayesiano Óptimo II

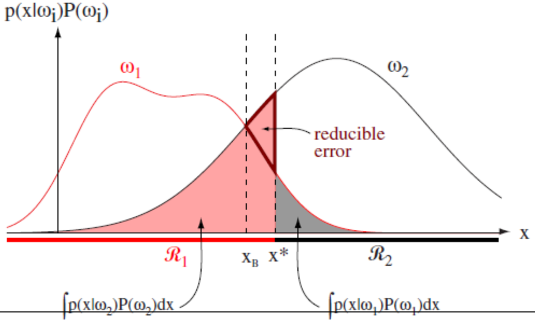
Alan Reyes-Figueroa

Introducción a la Ciencia de Datos (Aula 26) 12.abril.2021

Teorema (Optimalidad del clasificador Bayesiano)

Dada la información $(X, Y) \sim \mathbb{P}_{X,Y}$, sea $\widehat{y}$ la asignación del clasificador Bayesiano óptimo, y sea $\widetilde{y}$ cualquier otro clasificador. Entonces

$$\mathbb{P}_{\widehat{y}}(error \mid \mathbf{x}) \leq \mathbb{P}_{\widetilde{y}}(error \mid \mathbf{x}), \quad \forall \widetilde{y} : S \to \Omega.$$



Clasificador bayesiano óptimo

Si denotamos con L^* el error (promedio) del clasificador Bayesiano óptimo, y $\widetilde{y}_n$ es un clasificador basado en un conjunto finito de datos $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, con $L(\widetilde{y}_n) = \mathbb{E}\big(L(Y; \widetilde{y}_n(X))\big)$, se pueden demostrar las siguientes:

Propiedad

Si $\widetilde{y}_n$ es el clasificador 1-NN, entonces

$$L^* \leq \lim_{n\to\infty} \mathbb{E}\, L(\widetilde{y}_n) \leq 2L^*.$$

En general, Si $\widetilde{y}_n$ es el clasificador 1-NN para $M > 1$ clases, entonces

$$L^* \leq \lim_{n\to\infty} \mathbb{E}\, L(\widetilde{y}_n) \leq \frac{M}{M-1} L^*.$$

Propiedad

Si $n, k \to \infty$ son tales que $\frac{k}{n} \to 0$ y $\widetilde{y}_n$ es el clasificador k-NN, entonces

$$\lim_{n,k\to\infty} \mathbb{E}\, L(\widetilde{y}_n) = L^*.$$

- Las demostraciones de las propiedades anteriores se pueden ver en L. Devroye, L. Györfi, G. Lugosi (1996). *A Probabilistic Theory of Pattern Recognition.*

Clasificador bayesiano óptimo

Por otro lado: D.H. Wolpert, W.G. Macready (1997), *No Free Lunch Theorems for Optimization*, IEEE Transactions on Evolutionary Computation 1, 67.

Teorema (*No Free Lunch*)

Para n finito, sin ningún supuesto adicional sobre $\mathbb{P}$, ningún clasificador es mejor que otro.

Las prueba del *No Free Lunch Theorem* se pueden ver en

- S. Shalev-Shwartz, S. Ben-David (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge U. Press.
- `https://mlu.red/muse/52449366310.html`

Clasificador bayesiano ingenuo

Recordemos que en el clasificador Bayesiano óptimo

$$\widehat{y}(\mathbf{x}) = \mathbf{1}\left(\frac{\mathbb{P}(\mathbf{x} \mid Y = 1)}{\mathbb{P}(\mathbf{x} \mid Y = 0)} > \theta_\lambda\right), \quad \text{donde } {}_\lambda = \frac{(\lambda_{10} - \lambda_{00})\mathbb{P}(Y = 0)}{(\lambda_{01} - \lambda_{11})\mathbb{P}(Y = 1)}.$$

¿Cómo definir (estimar) $\mathbb{P}(X = \mathbf{x} \mid Y = y)$ en general?

- No evidente si X es de alta dimensión.
- El **clasificador Bayesiano ingenuo** (*Naive Bayes*) se basa en la simplificación (supuesto) que las componentes de $X = (X_1, \ldots, X_d) \mid Y$ son v.a. independientes:

$$\mathbb{P}(X = \mathbf{x} \mid Y = j) = \prod_{i=1}^{d} \mathbb{P}(X_i = x_i \mid Y = j).$$

- Llevamos 1 problema d-dimensional, a d problemas 1-dimensionales.

Clasificador bayesiano ingenuo

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Ejemplo

Ejemplo 1: Supongamos que Y es una v.a. discreta con $\mathbb{P}(Y=0)=\mathbb{P}(Y=1)=\frac{1}{2}$. Además

$$X \mid Y=0 \sim \mathcal{N}(2,1), \quad X \mid Y=1 \sim \mathcal{N}(5,1).$$

Construir el clasificador bayesiano óptimo, si

a) todos los errores con el mismo costo $\lambda = 1$.

b) $\lambda_{01}=4$, $\lambda_{10}=1$.

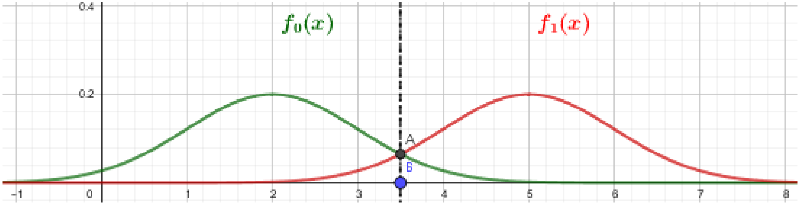
c) $\mathbb{P}(Y=0)=\frac{3}{5}$, $\mathbb{P}(Y=1)=\frac{2}{5}$.

Solución:

Tenemos

$$f_0(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}}, \quad f_1(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2}}.$$

Ejemplo



$$\lambda_{01} f_0(\mathbf{x})\,\mathbb{P}(Y=0) = \lambda_{10} f_1(\mathbf{x})\,\mathbb{P}(Y=1) \quad \Rightarrow \quad \tfrac{1}{2}(1)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}} = \tfrac{1}{2}(1)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-5)^2}{2}}$$

$$\Rightarrow \quad e^{-\frac{(x-2)^2}{2}} = e^{-\frac{(x-5)^2}{2}}$$

$$\Rightarrow \quad (x-2)^2 = (x-5)^2$$

$$\Rightarrow \quad x^2 - 4x + 4 = x^2 - 10x + 25$$

$$\Rightarrow \quad 6x = 21 \quad \Rightarrow \quad x = 3.5.$$

Ejemplo

Calcular el error:

Denotamos por $R_0 = (-\infty, 3.5)$, $R_1 = (3.5, \infty)$, las regiones de clasificación. Entonces

$$
\begin{aligned}
Error &= \int_{R_0} \mathbb{P}(X \mid Y = 1) + \int_{R_1} \mathbb{P}(X \mid Y = 0) \\
&= \int_{R_0} f_1(\mathbf{x})\, d\mathbf{x} + \int_{R_1} f_0(\mathbf{x})\, d\mathbf{x} = \int_{-\infty}^{3.5} f_1(\mathbf{x})\, d\mathbf{x} + \int_{3.5}^{\infty} f_1(\mathbf{x})\, d\mathbf{x} \\
&= \Phi^{-1}\left(\frac{3.5 - 5}{1}\right) + \left[1 - \Phi^{-1}\left(\frac{3.5 - 2}{1}\right)\right] \\
&= 2\Phi^{-1}(-1.5) = 2(0.06681) = 0.13362
\end{aligned}
$$

Así, nuestro clasificador de Bayes tiene un acierto del 86.6%.

Ejemplo

b) Como $\lambda_{01} = 1$ y $\lambda_{10} = 4$, aquí la ecuación a resulver resulta:

$$\lambda_{01} f_0(\mathbf{x})\,\mathbb{P}(Y=0) = \lambda_{10} f_1(\mathbf{x})\,\mathbb{P}(Y=1) \quad \Rightarrow \quad \tfrac{1}{2}(1)\tfrac{1}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}} = \tfrac{1}{4}(1).$$

de modo que $x 1.83$

Denotamos por $R_0 = (-\infty, 3.5)$, $R_1 = (3.5, \infty)$, las regiones de clasificación. Entonces

$$\begin{aligned}
Error &= \int_{R_0} \mathbb{P}(X \mid Y=1) + \int_{R_1} \mathbb{P}(X \mid Y=0) \\
&= \int_{R_0} f_1(\mathbf{x})\, d\mathbf{x} + \int_{R_1} f_0(\mathbf{x})\, d\mathbf{x} = \int_{-\infty}^{3.5} f_1(\mathbf{x})\, d\mathbf{x} + \int_{3.5}^{\infty} f_1(\mathbf{x})\, d\mathbf{x} \\
&= \Phi^{-1}\left(\frac{3.5-5}{1}\right) + \left[1 - \Phi^{-1}\left(\frac{3.5-2}{1}\right)\right] \\
&= 2\Phi^{-1}(-1.5) = 2(0.06681) = 0.13362
\end{aligned}$$

Ejemplo

Ejemplo 2: Supongamos que Y es una v.a. discreta con

$$\mathbb{P}(Y=1)=\tfrac{1}{2},\;\;\mathbb{P}(Y=2)=\tfrac{1}{4},\;\;\mathbb{P}(Y=3)=\tfrac{1}{4}.$$

Además,

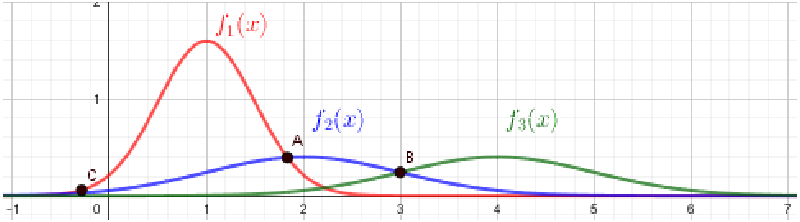
$$X\mid Y=1\sim\mathcal{N}(1,0.5^2),\;\;X\mid Y=2\sim\mathcal{N}(2,1),\;\;X\mid Y=3\sim\mathcal{N}(4,1).$$

Construir el clasificador bayesiano óptimo, con función de costo simétrica (todos los errores con el mismo costo $\lambda=1$).

Solución:
Tenemos

$$f_1(\mathbf{x})=\tfrac{2}{\sqrt{2\pi}}e^{-2(x-1)^2},\quad f_2(\mathbf{x})=\tfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}},\quad f_3(\mathbf{x})=\tfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-4)^2}{2}}.$$

Ejemplo



- B:

$$f_2(\mathbf{x})\,\mathbb{P}(Y=2) = f_3(\mathbf{x})\,\mathbb{P}(Y=3) \quad \Rightarrow \quad \frac{1}{4}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}} = \frac{1}{4}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-4)^2}{2}}$$

$$\Rightarrow \quad e^{-\frac{(x-2)^2}{2}} = e^{-\frac{(x-4)^2}{2}} \quad \Rightarrow \quad (x-2)^2 = (x-4)^2$$

$$\Rightarrow \quad x^2 - 4x + 4 = x^2 - 8x + 16$$

$$\Rightarrow \quad 4x = 12 \quad \Rightarrow \quad x = 3.$$

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Ejemplo

- A y C:

$$
\begin{aligned}
f_1(\mathbf{x})\,\mathbb{P}(Y=1) = f_2(\mathbf{x})\,\mathbb{P}(Y=2) \quad &\Rightarrow \quad \tfrac{1}{2}\cdot\tfrac{2}{\sqrt{2\pi}}e^{-2(x-1)^2} = \tfrac{1}{4}\cdot\tfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}}\\
&\Rightarrow \quad 4e^{-2(x-1)^2} = e^{-\frac{(x-2)^2}{2}}\\
&\Rightarrow \quad \log(4) - 2(x-1)^2 = -\tfrac{(x-2)^2}{2}\\
&\Rightarrow \quad 4(x-1)^2 - (x-2)^2 = 2\log(4)\\
&\Rightarrow \quad 3x^2 - 4x + 2\log(4) = 0\\
&\Rightarrow \quad x\tfrac{4\pm\sqrt{16+24\log(4)}}{6} \approx -0.5,\ 1.83.
\end{aligned}
$$

Entonces $\widehat{y}(x) = \begin{cases} 1, & \text{si } x \in (-0.5, 1.83);\\ 2, & \text{si } x \in (-\infty, -0.5) \cup (1.83, 3);\\ 3, & \text{si } x \in (3, \infty). \end{cases}$

Ejemplo

Calcular el error (ejercicio!).

Denotamos por $R_1 = (-0.5, 1.83)$, $R_2 = (-\infty, -0.5) \cup (1.83, 3)$, $R_3 = (3, \infty)$, las regiones de clasificación. Entonces

$$\begin{aligned}
Error &= \int_{R_1} \mathbb{P}(X \neq 1) + \int_{R_2} \mathbb{P}(X \neq 2) + \int_{R_3} \mathbb{P}(X \neq 3) \\
&= \int_{R_1} (f_2(\mathbf{x}) + f_3(\mathbf{x}))\, d\mathbf{x} + \int_{R_2} (f_1(\mathbf{x}) + f_3(\mathbf{x}))\, d\mathbf{x} + \int_{R_3} (f_1(\mathbf{x}) + f_2(\mathbf{x}))\, d\mathbf{x} \\
&= \int_{R_2 \cup R_3} f_1(\mathbf{x})\, d\mathbf{x} + \int_{R_1 \cup R_3} f_2(\mathbf{x})\, d\mathbf{x} + \int_{R_1 \cup R_2} f_3(\mathbf{x})\, d\mathbf{x} \\
&= \int_{-\infty}^{-0.5} f_1(\mathbf{x})\, d\mathbf{x} + \int_{1.83}^{\infty} f_1(\mathbf{x})\, d\mathbf{x} + \int_{-0.5}^{1.83} f_2(\mathbf{x})\, d\mathbf{x} + \int_{3}^{\infty} f_2(\mathbf{x})\, d\mathbf{x} + \int_{-\infty}^{3} f_3(\mathbf{x})\, d\mathbf{x} \\
&= ?
\end{aligned}$$

Ejemplo

En la práctica, es clasificador *Naive Bayes* es uno de los más utilizados (talvez no es el mejor, pero el clasificador más simple de construir).

Se basa en el hecho de estimar la distribución conjunta de (X, Y), a partir de la distribución conjunta empíricia (la tabla de datos).

Ejemplo 3: Considera la distrución conjunta de (X, Y) dada por

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	0.15	0.1	0.3
$Y = 1$	0.15	0.2	0.1

Ejemplo

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	0.15	0.1	0.3
$Y = 1$	0.15	0.2	0.1

Calcular el clasificador bayesiano óptimo si

a) la función de costo es la indicadora (falso positivo tiene el mismo costo 1 que un falso negativo).

b) si predecir mal un 1 (falso negativo) cuesta el doble que predecir mal un 0 (falso positivo).

Para ambos casos, calcula la probabilidad de cometer un error.

Ejemplo

<u>Solución</u>:
Vamos a tomar las probabilidades a priori de la tabla de datos:
$\mathbb{P}(Y=0)=0.55$, $\mathbb{P}(Y=1)=0.45$. Además los costos son $\lambda_{ij}=1-\delta_{ij}$.

- <u>$X=0$</u>:

$$\left.\begin{aligned} \lambda\mathbb{P}(Y=0\mid X=0) &= \lambda\frac{\mathbb{P}(X=0|Y=0)\,\mathbb{P}(Y=0)}{\mathbb{P}(X=0)} = 1\frac{0.15}{0.55}(0.55)=0.15 \\ \lambda\mathbb{P}(Y=1\mid X=0) &= \lambda\frac{\mathbb{P}(X=0|Y=1)\,\mathbb{P}(Y=1)}{\mathbb{P}(X=0)} = 1\frac{0.15}{0.45}(0.45)=0.15 \end{aligned}\right\} \Rightarrow \boxed{\widehat{y}(0)=0}.$$

- <u>$X=1$</u>:

$$\left.\begin{aligned} \lambda\mathbb{P}(Y=0\mid X=1) &= \lambda\frac{\mathbb{P}(X=1|Y=0)\,\mathbb{P}(Y=0)}{\mathbb{P}(X=1)} = 1\frac{0.1}{0.55}(0.55)=0.1 \\ \lambda\mathbb{P}(Y=1\mid X=1) &= \lambda\frac{\mathbb{P}(X=1|Y=1)\,\mathbb{P}(Y=1)}{\mathbb{P}(X=1)} = 1\frac{0.2}{0.45}(0.45)=0.2 \end{aligned}\right\} \Rightarrow \boxed{\widehat{y}(1)=1}.$$

Ejemplo

- $\underline{X = 2}$:

$$\left.\begin{array}{l} \lambda\mathbb{P}(Y = 0 \mid X = 2) = \lambda\frac{\mathbb{P}(X=2|Y=0)\,\mathbb{P}(Y=0)}{\mathbb{P}(X=2)} = 1\frac{0.3}{0.55}(0.55) = 0.3 \\ \lambda\mathbb{P}(Y = 1 \mid X = 2) = \lambda\frac{\mathbb{P}(X=2|Y=1)\,\mathbb{P}(Y=1)}{\mathbb{P}(X=2)} = 1\frac{0.1}{0.45}(0.45) = 0.1 \end{array}\right\} \Rightarrow \boxed{\widehat{y}(2) = 0}.$$

$$\text{Portanto, } \widehat{y}(x) = \begin{cases} 0, & \text{si } x = 0; \\ 1, & \text{si } x = 1; \\ 0, & \text{si } x = 2. \end{cases}$$

Calculamos el error:

$$\begin{aligned} Error &= 1 \cdot \mathbb{P}(Y = 1 \mid x = 0)\,\mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(Y = 0 \mid x = 1)\,\mathbb{P}(X = 1) + 1 \cdot \mathbb{P}(Y = 0 \mid x = 2)\,\mathbb{P}(X = \\ &= \tfrac{0.15}{0.3}(0.3) + \tfrac{0.1}{0.3}(0.3) + \tfrac{0.1}{0.4}(0.4) \\ &= 0.15 + 0.1 + 0.1 = \mathbf{0.35} \end{aligned}$$

Ejemplo

Parte (b): Aquí recordemos que $\lambda_{01} = 1$ (falso positivo), $\lambda_{10} = 2$ (falso negativo).

- $\underline{X = 0}$:

$$\left.\begin{aligned} \lambda\mathbb{P}(Y = 0 \mid X = 0) &= \lambda\tfrac{\mathbb{P}(X=0|Y=0)\,\mathbb{P}(Y=0)}{\mathbb{P}(X=0)} = 1\tfrac{0.15}{0.55}(0.55) = 0.15 \\ \lambda\mathbb{P}(Y = 1 \mid X = 0) &= \lambda\tfrac{\mathbb{P}(X=0|Y=1)\,\mathbb{P}(Y=1)}{\mathbb{P}(X=0)} = 2\tfrac{0.15}{0.45}(0.45) = 0.3 \end{aligned}\right\} \Rightarrow \boxed{\hat{y}(0) = 1}.$$

- $\underline{X = 1}$:

$$\left.\begin{aligned} \lambda\mathbb{P}(Y = 0 \mid X = 1) &= \lambda\tfrac{\mathbb{P}(X=1|Y=0)\,\mathbb{P}(Y=0)}{\mathbb{P}(X=1)} = 1\tfrac{0.1}{0.55}(0.55) = 0.1 \\ \lambda\mathbb{P}(Y = 1 \mid X = 1) &= \lambda\tfrac{\mathbb{P}(X=1|Y=1)\,\mathbb{P}(Y=1)}{\mathbb{P}(X=1)} = 2\tfrac{0.2}{0.45}(0.45) = 0.4 \end{aligned}\right\} \Rightarrow \boxed{\hat{y}(1) = 1}.$$

Ejemplo

- $\underline{X = 2}$:

$$\left.\begin{aligned} \lambda\mathbb{P}(Y=0 \mid X=2) &= \lambda\frac{\mathbb{P}(X{=}2|Y{=}0)\,\mathbb{P}(Y{=}0)}{\mathbb{P}(X{=}2)} = 1\,\frac{0.3}{0.55}(0.55) = 0.3 \\ \lambda\mathbb{P}(Y=1 \mid X=2) &= \lambda\frac{\mathbb{P}(X{=}2|Y{=}1)\,\mathbb{P}(Y{=}1)}{\mathbb{P}(X{=}2)} = 2\,\frac{0.1}{0.45}(0.45) = 0.2 \end{aligned}\right\} \Rightarrow \boxed{\widehat{y}(2) = 0}.$$

$$\text{Portanto, } \widehat{y}(x) = \begin{cases} 1, & \text{si } x = 0; \\ 1, & \text{si } x = 1; \\ 0, & \text{si } x = 2. \end{cases}$$

Calculamos el error:

$$\begin{aligned} \mathit{Error} &= 1\cdot\mathbb{P}(Y=0 \mid x=0)\,\mathbb{P}(X=0) + 1\cdot\mathbb{P}(Y=0 \mid x=1)\,\mathbb{P}(X=1) + 2\cdot\mathbb{P}(Y=1 \mid x=2)\,\mathbb{P}(X= \\ &= \frac{0.15}{0.3}(0.3) + \frac{0.1}{0.3}(0.3) + 2\frac{0.1}{0.4}(0.4) \\ &= 0.15 + 0.1 + 0.2 = \mathbf{0.45} \end{aligned}$$

Ejemplo

En el clasificador bayesiano, el caso multiclase se trabaja de forma análoga. Basta determinar aquella clase que tiene mayor probabilidad a *posteriori*:

$$
\begin{aligned}
\widehat{y}(\mathbf{x}) = i \quad & \Longleftrightarrow \quad i = \operatorname{argmax}_{1\leq j\leq k} \; \lambda_{j\ell}\, \mathbb{P}(Y = j \mid X = \mathbf{x}) \\
& \Longleftrightarrow \quad i = \operatorname{argmax}_{1\leq j\leq k} \; \lambda_{j\ell}\, \frac{\mathbb{P}(X = \mathbf{x} \mid Y = j)\, \mathbb{P}(Y = j)}{\mathbb{P}(X = \mathbf{x})} \\
& \Longleftrightarrow \quad i = \operatorname{argmax}_{1\leq j\leq k} \; \lambda_{j\ell}\, f_j(\mathbf{x})\, \pi_j.
\end{aligned}
$$

Así,

$$
\boxed{\widehat{y}(\mathbf{x}) = i \;\Leftrightarrow\; \lambda_{i\ell}\, f_i(\mathbf{x})\, \pi_i \geq \lambda_{j\ell}\, f_j(\mathbf{x})\, \pi_j, \;\; \forall j.}
$$

Ejemplo

Construir un clasificador de Bayes ingenuo para

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

y calcular la clasificación asociada a un vehículo rojo, SUV, doméstico.

Ejemplo

Queremos hallar la clasificación para un vehículo rojo, SUV, doméstico. Denotamos $X = (X_1, X_2, X_3)$, con $X_1 = color$, $X_2 = tipo$, $X_3 = origen$. Además $\mathbf{x} = (rojo, SUV, domestic)$.

Queremos comparar las probabilidades a *posteriori*

$$\mathbb{P}(X = \mathbf{x} \mid Y = 0)\, \mathbb{P}(y = 0) <> \mathbb{P}(X = \mathbf{x} \mid Y = 1)\, \mathbb{P}(y = 1).$$

Por independencia (naive Bayes), tenemos

$$\mathbb{P}(X = \mathbf{x} \mid Y = j) = \mathbb{P}(X_1 = \mathbf{x}_1 \mid Y = j)\, \mathbb{P}(X_2 = \mathbf{x}_2 \mid Y = j)\, \mathbb{P}(X_3 = \mathbf{x}_3 \mid Y = j)$$

- <u>Y=0</u>:

$$\mathbb{P}(X_1 = rojo|0) = \tfrac{2}{5},\ \ \mathbb{P}(X_2 = SUV|0) = \tfrac{3}{5},\ \ \mathbb{P}(X_3 = domestic|0) = \tfrac{3}{5}.$$

$$\Rightarrow\ \ f_0(\mathbf{x})\, \mathbb{P}(Y = 0) = \left(\tfrac{2}{5}\right)\left(\tfrac{3}{5}\right)\left(\tfrac{3}{5}\right)\left(\tfrac{5}{10}\right) = \tfrac{9}{125}.$$

Ejemplo

- <u>Y=1</u>:

$$\mathbb{P}(X_1 = rojo|1) = \tfrac{3}{5}, \;\; \mathbb{P}(X_2 = SUV|1) = \tfrac{1}{5}, \;\; \mathbb{P}(X_3 = domestic|1) = \tfrac{2}{5}.$$

$$\Rightarrow \;\; f_0(\mathbf{x})\,\mathbb{P}(Y=0) = \left(\tfrac{3}{5}\right)\left(\tfrac{1}{5}\right)\left(\tfrac{2}{5}\right)\left(\tfrac{5}{10}\right) = \tfrac{3}{125}.$$

Comparando ambas,

$$\left.\begin{array}{lr} Y=0: & f_0(\mathbf{x})\,\mathbb{P}(Y=0) = \frac{\mathbf{9}}{\mathbf{125}} \\ Y=1: & f_1(\mathbf{x})\,\mathbb{P}(Y=1) = \frac{3}{125} \end{array}\right\} \Rightarrow \;\; \widehat{y}(\mathbf{x}) = 0.$$

De ahí que nuestro clasificador Naive Bayes asigna los vehículos de tipo $\mathbf{x} =$ (rojo, SUV, doméstico) a la clase $y = 0$.