Clasificador de Bayes naive.

Supone que las variables predictivas se comporten de manera independiente y en caso que sean continuas provienen de densidades normales.

$$\boxed{P(Y=1|X=x) = \frac{P(Y=1)\,P(X=x|Y=1)}{P(X=x)}}$$

Ejemplo: ej 1 Práctico

Training S.

x_1	a	a	b	a	a	b	b	b
x_2	b	a	a	a	a	b	b	b
y	1	1	1	1	-1	-1	-1	-1

Test S

a	a	b	b
a	b	a	b
?	?	?	?

$$P(Y=1|X=(a,a)) = \frac{P(Y=1)\,P(X=(a,a)|Y=1)}{P(X=(a,a))}$$

indep. condicional.

$$\overset{\downarrow}{=} \frac{1/2\;P(x_1=a|Y=1)\,P(x_2=a|Y=1)}{\underbrace{P(X=(a,a)|Y=1)\,P(Y=1)+P(X=(a,a)|Y=-1)\,P(Y=1)}_{P(X=(a,a))}}$$

uso la formula de proba total.

$$= \frac{1/2 \cdot 3/4 \cdot 3/4}{P(x_1=a|Y=1)\,P(x_2=a|Y=1)\,P(Y=1)+P(x_1=a|Y=-1)\,P(x_2=a|Y=-1)\,P(Y=-1)}$$

Naive Bayes. (Ejercicio 1 Práctico Bin. Clas)

$$P(B|A) = \frac{P(B)}{P(A)} P(A|B).$$

TS

X_1	a	a	b	a	a	b	b	b
X_2	b	a	a	a	a	b	b	b
Y	1	1	1	1	-1	-1	-1	-1

Tests.

	a	a	b	b
	a	b	a	b
	?	?	?	?
Real	1	-1	1	1

$$P(Y=1 | X=(a,a)) = \frac{P(Y=1)}{P(X=(a,a))} P(X=(a,a) | Y=1)$$

$$= \frac{1/2 \; P(X_1=a | Y=1)\, P(X_2=a | Y=1)}{P(X=(a,a)|Y=1)\,P(Y=1) + P(X=(a,a)|Y=-1)\,P(Y=1)}$$

$$= \frac{1/2 \cdot 3/4 \cdot 3/4}{P(X_1=a | Y=1)P(X_2=a|Y=1)P(Y=1) + P(X_1=a|Y=-1)P(X_2=a|Y=-1)P(Y=-1)}$$

$$= \frac{1/2 \cdot 3/4 \cdot 3/4}{3/4 \cdot 3/4 \cdot 1/2 + 1/4 \cdot 1/4 \cdot 1/2} = \frac{9/32}{9/32 + 1/32} = \frac{9/32}{10/32} = \frac{9}{10} = 0.9$$

$$\boxed{X=(a,a) \rightsquigarrow 1}$$

$$P(Y=-1 | X=(a,a)) = 0.1$$

$$P(Y=1 | X=(a,b)) = \frac{P(Y=1)}{P(X=(a,b))} P(X=(a,b) | Y=1)$$

$$= \frac{1/2 \; P(X_1=a|Y=1)\,P(X_2=b|Y=1)}{P(X_1=a|Y=1)P(X_2=b|Y=1)\,P(Y=1) + P(X_1=a|Y=-1)P(X_2=b|Y=-1)P(Y=-1)}$$

$$= \frac{1/2 \cdot 3/4 \cdot 1/4}{3/4 \cdot 1/4 \cdot 1/2 + 1/4 \cdot 3/4 \cdot 1/2} = \frac{3/32}{3/32 + 3/32} = 0.5$$

$$X=(a,b) \rightsquigarrow -1$$

Clasificador Bayesiano. (Wikipedia)

Entrenamiento [editar]

Entrenamiento previo.

sexo	altura (pies)	peso (lbs)	talla del pie (inches)
hombre	6	180	12
hombre	5.92 (5'11")	190	11
hombre	5.58 (5'7")	170	12
hombre	5.92 (5'11")	165	10
mujer	5	100	6
mujer	5.5 (5'6")	150	8
mujer	5.42 (5'5")	130	7
mujer	5.75 (5'9")	150	9

Haciendo una distribución Gaussiana extraemos los datos y obtenemos la media y la varianza de cada característica.

sexo	media (altura)	varianza (altura)	media (peso)	varianza (peso)	media (talla del pie)	varianza (talla del pie)
hombre	5.855	0.035033	176.25	122.92000	11.25	0.91667
mujer	5.4175	0.097225	132.5	558.33000	7.5	1.66670

En este caso nos encontramos en una distribución equiprobable, es decir que tienen la misma probabilidad. P(hombre)=0.5 y P(mujer)=0.5.

Testing [editar]

Ahora recibimos unos datos para ser clasificado como hombre o mujer

sex	altura (pies)	peso (lbs)	número de pie(inches)
muestra	6	130	8

Clasificador Bayesiano. (Wikipedia)

Entrenamiento [editar]

Entrenamiento previo.

sexo	altura (pies)	peso (lbs)	talla del pie (inches)
hombre	6	180	12
hombre	5.92 (5'11")	190	11
hombre	5.58 (5'7")	170	12
hombre	5.92 (5'11")	165	10
mujer	5	100	6
mujer	5.5 (5'6")	150	8
mujer	5.42 (5'5")	130	7
mujer	5.75 (5'9")	150	9

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$$

Haciendo una distribución Gaussiana extraemos los datos y obtenemos la media y la varianza de cada característica.

sexo	media (altura) $\hat{\mu}_{alt}$	varianza (altura) $\hat{\sigma}^2_{alt}$	media (peso) $\hat{\mu}_{peso}$	varianza (peso) $\hat{\sigma}^2_{peso}$	media (talla del pie) $\hat{\mu}_{pie}$	varianza (talla del pie) $\hat{\sigma}^2_{pie}$
hombre	5.855	0.035033	176.25	122.92000	11.25	0.91667
mujer	5.4175	0.097225	132.5	558.33000	7.5	1.66670

En este caso nos encontramos en una distribución equiprobable, es decir que tienen la misma probabilidad. $\underbrace{P(\text{hombre})=0.5}_{\pi_1}$ y $\underbrace{P(\text{mujer})=0.5}_{\pi_2}$.

Testing [editar]

Ahora recibimos unos datos para ser clasificado como hombre o mujer

sex	altura (pies)	peso (lbs)	número de pie(inches)
muestra	6	130	8

x_0. ¡nueva observación!

$$\overset{?}{P}(H \mid x_0) = \frac{P(H)}{P(x_0)} \cdot P(x_0 \mid H)$$

Supongo estoy trabajando con densidades normales

$P(H) = P(M) = 0.5$

Densidad normal

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Comparo $P(H)\,P(x_0|H)$ contra $P(M)\,P(x_0|M)$

pues $P(H|x_0) = \dfrac{P(H)\,P(x_0|H)}{P(x_0)}$ y $P(M|x_0) = \dfrac{P(M)\,P(x_0|M)}{P(x_0)}$

$$P(x_0|H) = P(alt=6,\ peso=130,\ talla=8\,|H)$$

indep. condicional

$$= P(alt=6|H) \cdot P(peso=130|H) \cdot P(talla=8|H)$$

$$= \underbrace{f_{alt,\,hombre}(6)}_{(1)}\ \underbrace{f_{peso,\,hombre}(130)}_{(2)}\ \underbrace{f_{talla,\,hombre}(8)}_{(3)}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{0.035}}\, e^{-\frac{1}{2}\left(\frac{6-5{,}855}{\sqrt{0.035}}\right)^2}}_{(1)} \cdot \underbrace{\quad}_{(2)} \cdot \underbrace{\quad}_{(3)}$$

$$= 1{,}5789 \cdot 5{,}98 \cdot 10^{-6} \cdot 1.311 \cdot 10^{-3}$$

$$P(x_0) = P(x_0|H)\,P(H) + P(x_0|M)\,P(M)$$

$$= P(alt=183|H)\,P(peso=59|H)\,P(talla=20|H)$$

$$+ P(alt=183|M)\,P(peso=59|M)\,P(talla=20|M)$$

(no es necesario

$$\Rightarrow P(H|x_0) = \frac{\overset{P(H)}{0.5}\ \cdot}{P(x_0)} = \frac{6.1984 \cdot 10^{-9}}{P(x_0)}$$

$$P(M|x_0) = \frac{5.3778 \cdot 10^{-4}}{P(x_0)}$$

$\Longrightarrow$ clasifico como M.