CLASIFICADOR BAYESIANO INGENUO
(NAIVE BAYES)

Teorema de Bayes:

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

- Demostración:

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

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

Causas: $A_1, \ldots, A_n$

Suceso: B

VEROSIMILITUD

Sucede B: ¿Prob. de que la causa sea A_k? A PRIORI

A POSTERIORI

$$\boxed{P(A_k|B)} = \frac{P(B|A_k)\,P(A_k)}{P(B)} = \frac{\boxed{P(B|A_k)}\,\boxed{P(A_k)}}{\boxed{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)}}$$

EVIDENCIA

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

EJEMPLO

5. En una asignatura universitaria de primero asisten a clase 100 de los 150 alumnos.
Se sabe que aprueban
el 90% de los alumnos que asisten a clase y el 30% de los que no asisten.
Se elige un alumno al azar. Calcular:
a. La probabilidad de que haya suspendido.
b. Si se sabe que el alumno ha suspendido, la probabilidad de que haya asistido a clase.

Causas: $A_1 \leftarrow$ Asiste a clase

$A_2 \leftarrow$ No a clase

Suceso: $B \leftarrow$ Suspende

(a)
$$P(B) = \underbrace{P(B|A_1)}_{1-0{,}9=0{,}1} \cdot \underbrace{P(A_1)}_{\frac{100}{150}} + \underbrace{P(B|A_2)}_{1-0{,}3=0{,}6} \cdot \overbrace{P(A_2)}^{50/150}$$

$$= \frac{10}{150} + \frac{30}{150} = \frac{40}{150} = \frac{4}{15}$$

(b)
$$P(A_1|B) = \frac{P(B|A_1)\cdot P(A_1)}{P(B)} = \frac{1/15}{4/15} = \frac{1}{4}$$

• Clasificador bayesiano

Variable categórica $\rightarrow$ $\boxed{\text{Target}} \rightarrow \boxed{y} = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{pmatrix}$ CATEGORÍAS $C_1, \ldots, C_N$

Training data: $X = \begin{pmatrix} x_1^{(1)} & \cdots & x_m^{(1)} \\ \vdots & & \\ x_1^{(N)} & \cdots & x_m^{(N)} \end{pmatrix} = \begin{pmatrix} \vec{x}^{(1)} \\ \vdots \\ \vec{x}^{(N)} \end{pmatrix}$ [Ej. Va a clase, No va a clase]

Dada una nueva instancia: $\vec{x} = (x_1, \ldots, x_m) \rightsquigarrow$ ¿y?

¿Prob. de que $\vec{x}$ esté en la categoría C_K?

¿$P(y = C_K)$?

$$\boxed{P(C_K|\vec{x})}$$

(TEST DATA) A POSTERIORI

$$\boxed{P(C_K|\vec{x})} = \frac{\boxed{P(\vec{x}|C_K)} \cdot \boxed{P(C_K)}}{\boxed{P(\vec{x})}}$$

A PRIORI (TRAINING DATA)

EVIDENCIA

VEROSIMILITUD (LIKELIHOOD)

$$P(C_K) = \frac{\#\text{ datos en } C_K}{\#\text{ datos total}}$$

$\#\text{ datos total} = N$

Quiero hallar el K que maximiza $P(C_K|\vec{x})$

[Predecir C_K para la nueva instancia $\vec{x}$]

→ $P(\vec{x})$ no afecta al máximo ←

$\hat{y} = C_{\boxed{\hat{K}}}$

Quiero predecir la categoría

$$\hat{K} = \underset{K}{\text{indmax}}\ P(C_K|\vec{x})$$

$\equiv$ El K tal que $P(C_K|\vec{x})$ es máximo

$$\hat{K} = \underset{K}{\text{indmax}}\ P(C_K|\vec{x}) = \boxed{\underset{K}{\text{indmax}}\ P(\vec{x}|C_K) \cdot P(C_K)}$$

Esta ya la tengo

Esta no la tengo

• Clasificador bayesiano ingenuo

(Naive assumption)

Suposición ingenua

$\vec{x} = (x_1, \ldots, x_n)$

Supongo que $P(x_1|C_K), \ldots, P(x_n|C_K)$ son variables independientes,

es decir: $P(\vec{x}|C_K) = P(x_1|C_K) \times P(x_2|C_K) \times \ldots \times P(x_n|C_K)$

[Asumo que las features son independientes]

Suponemos que conocemos las distribuciones de las

$P(x_i|C_K)$ a priori:

Por ejemplo, las estimamos como distribuciones normales: $P(x_i|C_K) \sim N(x_i|\mu_K^i, \sigma_K^i)$

Media de x_i en C_K

Varianza de x_i en C_K

$$\boxed{\hat{K} = \operatorname{indmax}_K \left[N(x_1|\mu_K^1,\sigma_K^1)\dots N(x_n|\mu_K^n,\sigma_K^n)\, \frac{\#C_K}{N} \right]}$$

El clasificador bayesiano ingenuo, asigna a una nueva columna (x1,...,xn) la categoría que maximice el número:

$$N(x_1|\mu_K^1,\sigma_K^1) \times \dots \times N(x_n|\mu_K^n,\sigma_K^n)\, \frac{\#C_K}{N}$$

$$\boxed{N(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

Recordad