# T4-incertidumbre.pdf

Olmar_eps

Inteligencia Artificial

3º Grado en Ingeniería Informática

Escuela Politécnica Superior
Universidad Autónoma de Madrid

# Tema 4: INCERTIDUMBRE

## 4.1. FORMALIZACIÓN de INCERTIDUMBRE mediante PROBABILIDAD

- Ley de Probabilidad Total:

$$P(B) = \sum_{i=1}^{n} P(B, A=a_i) = \sum_{i=1}^{n} P(B \mid A=a_i)\, P(A=a_i)$$

- Probabilidad Condicional

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

- Regla de la Suma

$$P(A) + P(\bar{A}) = 1 \qquad P(A \mid I) + P(\bar{A} \mid I) = 1$$

- Regla del producto

$$P(A,B) = P(A \mid B)\, P(B) = P(B \mid A)\, P(A) \qquad \boxed{P(A,B \mid I) = P(B \mid A, I)\, P(A \mid I)}$$

- Regla de Bayes

verosimilitud hipótesis dado los datos

prob. PRIORI

$$P(A \mid B) = \frac{\overbrace{P(B \mid A)}^{\text{verosimilitud}} \cdot P(A)}{\underbrace{P(B)}_{\text{evidencia de los datos}}} \qquad \boxed{P(A \mid B, I) = \frac{P(B \mid A, I)\, P(A \mid I)}{P(B \mid I)}}$$

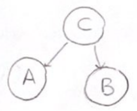
prob. POSTERIORI

- Independencia

A y B independientes si $P(A,B) = P(A \mid B)\, P(B) = P(A) \cdot P(B)$

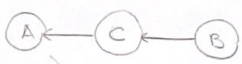
A es condicionalmente independiente de B dado C si:

$$P(A \mid B, C) = P(A \mid C) \quad \equiv \quad P(A, B \mid C) = P(A \mid C)\, P(B \mid C)$$




A condic indep de B

ó




condic indep

## 4.2. TEOREMA de BAYES en IA

→ PRIOR: probabilidad a priori de la hipótesis H.

Probabilidad de la hipótesis, <u>antes</u> de observar los datos

$$P(H) \longrightarrow \sum_H P(H) = 1$$

→ POSTERIOR: probabilidad a posteriori de la hipótesis H.

Probabilidad de la hipótesis, <u>después</u> de observar los datos.

$$P(H|D) \longrightarrow \sum_H P(H|D) = 1$$

Difícil de estimar (condicionado a D)

→ VEROSIMILITUD: de la hipótesis dado los datos.

$$P(D|H)$$

Más confiables que las posteriores

→ EVIDENCIA: de los datos.

Es independiente de la hipótesis y funciona como un factor de normalización.

$$P(D) = \sum_H P(D|H)\,P(H) \rightsquigarrow \text{cte de normalización}$$

### INFERENCIA a partir de los DATOS (seleccionando hipótesis)

↳ PRIOR: $H^*_{prior} = \underset{H}{\operatorname{argmax}}\, P(H)$ (ignora información de D)

↳ ML: $H^*_{ML} = \underset{H}{\operatorname{argmax}}\, P(D|H)$ (no usa información de los prioris ≡ prior uniforme)

máx. veros. → asigna la clase que maxim. la veros.

miro tabla y filtro por (10)

↳ MAP: $H^*_{MAP} = \underset{H}{\operatorname{argmax}}\, P(H|D)$ → Bayes es óptimo ⇝ minimiza el error

máx. posterior → asigna la clase que maxim. prob. post.

↳ INFERENCIA BAYESIANA: $\mathbb{E}(f(H)) = \sum_H f(H)\,P(H|D)$ → promedia sobre hipótesis con probabilidades

* Si las estimaciones de posteriores son fiables, MAP es mejor

* Seleccionar ML puede ser ventajoso cuando los prioris son desconocidos.

* $H^*_{ML} = H^*_{MAP}$ si los prioris de todas las hipótesis son iguales

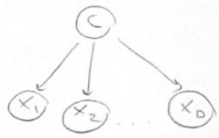
## NAÏVE BAYES

NB asume que los atributos son independientes condicionados a la clase.

$$P(X|C) = P(X_1|C)\,P(X_2|C)\dots P(X_D|C)$$

Teorema de Bayes

$$P(C|X) = \frac{P(X|C)\cdot P(C)}{P(X)} = \frac{P(X_1|C)P(X_2|C)\dots P(X_D|C)}{P(X)}$$




Ⓥ: Entrenamiento y precisión rápidos

Ⓓ: Las estimaciones de probabilidad no son confiables

## ESTIMACIÓN de PROBABILIDADES

Las estimaciones de frecuencia de probabilidades pueden no ser fiables, especialmente cuando las muestras son pequeñas (términos a 0)

⇒ ESTIMADOR de LAPLACE

- Evita estimaciones nulas para las probabilidades
- Estimaciones más robustas
- Asintóticamente pequeño

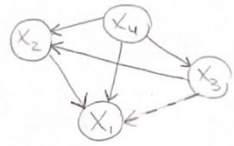
$$P_i = \frac{n_i + \frac{\mu}{k}}{n_{total} + \mu} \xrightarrow{\text{típicamente } \mu = k} P_i = \frac{n_i + 1}{n_{total} + \mu}$$

ejemplo ficticio

## 4.3. RED BAYESIANA

$$P(\text{nodo} \mid \text{padres}(\text{nodo})) \Rightarrow P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{padres}(X_i))$$

ej:




$$P(X_1|X_2,X_3,X_4)\cdot P(X_2|X_3,X_4)\cdot P(X_3|X_4)\cdot P(X_4)$$

- $X_2 \to X_1$ significa que $X_2$ está entre las variables en las que $X_1$ es condicionada.
- Si no pongo relación entre dos nodos ($X_3 \dashrightarrow X_1$) significa que $X_3$ es cond. indep de $X_1$, dados $X_2$ y $X_4$.