

4. Incertidumbre en IA



Incertidumbre en IA

- ☐ Formalización de incertidumbre mediante probabilidades
- ☐ Teorema de Bayes en IA
- ☐ Redes Bayesianas

Lecturas recomendadas:

- CAPÍTULO 1 del libro de Bishop
- CAPÍTULOS 1, 2 del libro de Jaynes

4.1 Formalización de incertidumbre mediante probabilidades

Las probabilidades, entendidas como la confianza en la verosimilitud de una proposición, pueden usarse para extender de manera coherente el razonamiento lógico a situaciones con incertidumbre.

Cox "Probability, Frequency, and Reasonable Expectation," Am. Jour. Phys. 14, 1–13, (1946)

❑ El **grado de plausibilidad** de una aserción puede ser representada por un número real que refleja la información disponible relacionada con la aserción.

❑ Correspondencia cualitativa con el **sentido común**.

“[...]la théorie des probabilités n'est au fond, que le bon sens réduit au calcul [...]” (Pierre-Simon Laplace “Essai philosophique sur les probabilités “, p. 275 Paris: Bachelier, Courcier, 1825)

❑ **Consistencia**. Si el grado de plausibilidad de una aserción puede ser derivado de diferentes formas, los valores computados deben ser iguales.

Probabilidades

- ❑ La probabilidad puede interpretarse como una medida de:
 - ❑ Proporción de veces que algo es cierto
 - ❑ 20 estudiantes aprobaron el examen de 22
 - ❑ un fenómeno físico puede medirse experimentalmente
 - ❑ Grado de creer algo
 - ❑ “Creo que el Real Madrid ganará la Liga con un 80%”
 - ❑ puede variar en las personas o sistemas inteligentes
- ❑ El cálculo de probabilidad no depende de la interpretación
 - ❑ Rango de probabilidades en $[0.0..1.0]$
 - ❑ Probabilidad = 0: falso
 - ❑ Probabilidad = 1: verdadero

Probabilidades y causalidad

❑ Las probabilidades representan conexiones lógicas, no conexiones causales.

❑ $A \Rightarrow B$ no debería ser entendido como “A es la causa física de B”

❑ Por equivalencia $\neg B \Rightarrow \neg A$ y “ $\neg B$ es la causa física de $\neg A$ ” (?)

Ej. $\neg \text{BATERÍA_OK} \Rightarrow \neg \text{FUNCIONA}$ (causal?)

“El dispositivo no funciona porque la batería no está OK”

$\text{FUNCIONA} \Rightarrow \text{BATERÍA_OK}$ (NO CAUSAL)

El funcionamiento del aparato no es la causa física de que la batería esté OK

Ej. Las nubes son la causa física de la lluvia.

Sin embargo, $\text{Nubes} \Rightarrow \text{Lluvia}$ es incorrecto.

La aserción correcta es $\text{Lluvia} \Rightarrow \text{Nubes}$, que no puede ser entendida como “la lluvia es la causa física de las nubes”

Probabilidades y causalidad

❑ Experimento con extracción de una urna

❑ Experimento 1: Urna con 1 bola roja y 5 bolas negras.

- ❑ La probabilidad de extraer una bola roja es $1/6$.

❑ Experimento 2: Urna con 1 bola roja y 5 bolas negras.

- ❑ Se extrae una bola roja de la urna y no se devuelve a ella.

En una segunda extracción se extrae una bola.

Dado que sólo hay una bola roja en la urna, la probabilidad de haber observado una bola roja en la segunda extracción es 0.

La probabilidad del resultado de la segunda extracción depende del resultado de la primera.

Sin embargo, la segunda extracción no puede afectar causalmente la primera.

Uso de Probabilidades en IA

- ❑ Tareas típicas: toma de decisiones, clasificación, predicción, . . .
 - ❑ ¿Qué es verdad? un fenómeno físico
 - ❑ uso de la lógica clásica: satisfacción proposicional, sistemas de producción, camino más corto, ajedrez, etc
 - ❑ vs. ¿Qué es más probable?
 - ❑ Uso de probabilidades: redes bayesianas, predicción de secuencias (reconocimiento de voz), clasificación (del lenguaje), pronóstico del tiempo, videojuegos
 - ❑ ¿Qué pasa si el modelo seleccionado está mal?
 - ❑ En la lógica clásica:
 - ❑ modelo incompleto \rightarrow ok
 - ❑ modelo incorrecto \rightarrow problema
 - ❑ En probabilidades
 - ❑ Por lo general, es más interesante conocer la relación entre las probabilidades que los números exactos: $P(e) > P(e_j)$?
 - ❑ podría ser más robusto

Variables aleatorias

❑ Lógica proposicional

- ❑ describimos los estados como conjuntos de variables booleanas: p, q, r
- ❑ una interpretación es una asignación de verdad a esas variables: $p = V, q = F, r = V$

❑ Teoría de probabilidad

- ❑ usamos un conjunto de variables aleatorias que pueden tomar valores en un dominio dado:
 - ❑ un dado: $X \in \{1, 2, 3, 4, 5, 6\}$
 - ❑ dos dados: $X \in \{1, 2, 3, 4, 5, 6\}, Y \in \{1, 2, 3, 4, 5, 6\}$
- ❑ el valor asociado a una variable aleatoria es desconocido
- ❑ podemos asignar una probabilidad a cada valor
 - ❑ dados: $P(X = 1) = 1/6, \dots, P(X = 6) = 1/6$
- ❑ estas probabilidades definen una distribución de probabilidad

Ejemplo

- Dado un robot en una cuadrícula de 100×100 con una orientación dada. Definir sus variables aleatorias, dominios y distribución de probabilidad:
 - Variables aleatorias
 - $X \in \{0, \dots, 99\}$, $Y \in \{0, \dots, 99\}$, $\theta \in \{0, \dots, 359\}$

Probabilidad a priori

- La distribución de probabilidad de una variable aleatoria generalmente se representa como un (vector)
 - Ejemplo: $P(X) = (P(X = 0), \dots, P(X = 99))$
- La distribución de probabilidad conjunta es la distribución de varias variables. E.g. $P(X, Y)$, $P(X, Y, \theta)$

$$\begin{aligned} P(X, Y) &= (P(X = 0, Y = 0), P(X = 0, Y = 1), \dots, P(X = 99, Y = 99)) \\ &= \left(\frac{1}{10000}, \frac{1}{10000}, \dots, \frac{1}{10000} \right) \end{aligned}$$

- Esta distribución es "a priori" o incondicional, ya que no depende de ninguna condición.

Ejemplo

- Dado un robot en una cuadrícula de 100×100 con una orientación dada
- Definir sus variables aleatorias, dominios y distribución de probabilidad.

- Variables aleatorias

$$X \in \{0, \dots, 99\}, Y \in \{0, \dots, 99\}, \theta \in \{0, \dots, 359\}$$

- Distribución de probabilidad (posición desconocida):

$$P(X = 0, Y = 0) = P(X = 0, Y = 1) = \dots =$$

$$P(X = 0, Y = 99) = P(X = 1, Y = 0) = \dots =$$

$$P(X = 99, Y = 99) = \frac{1}{100 \times 100}$$

- Distribución de probabilidad (posición y orientación desconocidas):

- $P(X = 0, Y = 0, \theta = 0) = \dots$

- $P(X = 100, Y = 100, \theta = 360) = \frac{1}{100 \times 100 \times 360}$

Ley de probabilidad total (total probability)

□ Dado un conjunto de eventos A_i disjuntos por pares, de manera que su unión es el espacio muestral completo y otro evento B :

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

□ Así, si tenemos una variable aleatoria A con posibles valores disjuntos a_1, \dots, a_n y un evento B :

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

□ Ejemplo

$$\square P(X = 0) = \sum_{i=1}^{99} P(X = 0, Y = i) = P(X = 0, Y = 0) + P(X = 0, Y = 1) + \dots + P(X = 0, Y = 99) = \frac{1}{10000} + \frac{1}{10000} + \dots + \frac{1}{10000} = \frac{100}{10000} = \frac{1}{100}$$

Probabilidad condicional

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

- $P(A | B)$ puede interpretarse como la probabilidad actualizada de A, una vez que se ha observado B

□ Ejemplos

- $P(X = 0)$?
- $P(X = 0 | X < 10)$?
- $P(X = 0 | Y = 0)$?
- $0 \leq P(A | I) \leq 1$
- $P(\text{True} | I) = 1, P(\text{False} | I) = 0$

□ Regla de la suma:

- $P(A) + P(\neg A) = P(A) + P(\bar{A}) = 1$

Probabilidad condicional

- Probabilidad Condicional

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \quad \text{si } P(B) \neq 0$$

- Regla del producto

$$P(A, B) = P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$

- Regla de Bayes

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

- A veces obtener $P(A|B)$ es más fácil que $P(B|A)$

$$P(\text{Causa} | \text{Efecto}) = \frac{P(\text{Efecto} | \text{Causa})P(\text{Causa})}{P(\text{Efecto})}$$

Normalmente es más fácil preguntar a un experto $P(\text{Efecto} | \text{Causa})$ que $P(\text{Causa} | \text{Efecto})$

Independencia

- ❑ A y B son independientes si:
 - ❑ $P(A|B) = P(A)$;
 - ❑ $P(B|A) = P(B)$; o
 - ❑ $P(A,B) = P(A|B)P(B) = P(A)P(B)$
- ❑ Ej. $P(\text{RobotX}, \text{Orientación}, \text{Dado}) = P(\text{RobotX}) P(\text{Orientación}) P(\text{Dado})$
 - ❑ Reducción en el tamaño de distribución:
 - ❑ $100 \times 360 \times 6 = 216000 \sim 100 + 360 + 6 = 466$
- ❑ Una descripción más pequeña implica algoritmos más eficientes
- ❑ Menos datos (probabilidades) a especificar

Incertidumbre en IA

- Formalización de incertidumbre mediante probabilidades
- Teorema de Bayes en IA
- Redes Bayesianas

4.2 Teorema de Bayes en IA

Teorema de Bayes

$$P(H | D) = \frac{P(D|H)P(H)}{P(D)}$$

Teorema de la probabilidad total

Si tenemos las variables H_1, \dots, H_n y un evento D :

$$P(D) = \sum_{i=1}^n P(D|H_i) P(H_i)$$

H: Hipótesis

D: Datos

$P(H)$: **Probabilidad a priori** de la hipótesis H .
Probabilidad de la hipótesis, antes de observar los datos.

$P(D | H)$: **Verosimilitud** de la hipótesis dado los datos.

$P(D)$: **Evidencia** de los datos.
Es independiente de la hipótesis y funciona como un factor de normalización.

$P(H | D)$: **Probabilidad a posteriori** de la hipótesis.
Probabilidad de la hipótesis, después de observar los datos.

Inferencia

❑ Calcular las probabilidades de eventos **e** dada alguna evidencia **o**: $P(e | o)$

❑ Ejemplos

- ❑ Calcular la distribución a posteriori dada una evidencia
- ❑ Elegir una acción para lograr una alta recompensa y alguna evidencia
- ❑ Toma de decisiones con utilidad óptima
- ❑ Clasificación
- ❑ Diagnóstico

Inferencia

Calcula la probabilidad a posteriori dado $P(X|o)$

x0 x1 x2 x3 x4 x5 x6 x7



❑ ¿Cuál es la distribución de probabilidad de la posición del robot (X) dado que observó una puerta? ($o = \text{door}$), $P(X | o = \text{door})$?

$$\square P(X | o = \text{door}) = (0, 0, \frac{1}{2}, 0, 0, 0, 0, \frac{1}{2})$$

$$\square P(X=0 | o = \text{door}) = \frac{P(\text{door} | X=0)P(x=0)}{P(\text{door})} = \frac{0 \times \frac{1}{8}}{\frac{2}{8}} = 0$$

$$\square P(X=2 | o = \text{door}) = \frac{P(\text{door} | X=2)P(x=2)}{P(\text{door})} = \frac{1 \times \frac{1}{8}}{\frac{2}{8}} = 0$$

Inferencia

❑ Clasificación

- ❑ dadas algunas observaciones a qué clase pertenecen?
- ❑ compara $P(\text{Clase} = 1 \mid o)$ versus $P(\text{Clase} = 2 \mid o)$
- ❑ ejemplos:
 - ❑ dado algunos datos del cliente (dinero promedio en el banco, ubicación de la casa, ganancias mensuales), determinar si devolverá correctamente una hipoteca (Clase = 1) o no (Clase = 2)
 - ❑ dada alguna imagen (número de píxeles de una luminosidad dada, número de líneas), determinar si pertenece a un gato (Clase = gato), un perro (Clase = perro), o algo diferente (Clase otro)

❑ Computación (Naïve Bayes)

- ❑ $\text{Clase} = \arg \max_{c \in \text{Clases}} P(\text{Clase} = c \mid o)$

❑ Diagnósis

- ❑ probabilidad de una enfermedad $I = 1$ dados los resultados del análisis o , $P(I = 1 \mid o)$

Inferencia a partir de los datos

❑ Máxima verosimilitud (ML):

Selecciona la hipótesis que maximiza la verosimilitud de la hipótesis dados los datos

$$H^* = \arg \max_H P(D | H)$$

❑ ML no usa información de los prioris (equivale a asumir un priori uniforme, $P(H) = \text{constante}$).

❑ Máxima probabilidad a posteriori (MAP)

Selecciona la hipótesis que maximiza la probabilidad a posteriori

$$H^* = \arg \max_H P(H | D) = \arg \max_H \frac{P(D|H) P(H)}{P(D)} = \arg \max_H P(D | H) P(H)$$

❑ Inferencia Bayesiana.

Promediar sobre todas las hipótesis con probabilidades

Inferencia Bayesiana

- ❑ La teoría de decisión bayesiana se basa en dos supuestos
 - ❑ El problema de decisión se puede describir en términos probabilísticos
 - ❑ Todas las probabilidades del problema son conocidas o al menos pueden estimarse
- ❑ Las decisiones se toman en función de los datos observados
- ❑ Notación
 - ❑ Conjunto de clases: $C = \{c_1, c_2, \dots, c_m\}$
 - ❑ Conjunto de atributos: $A = \{a_1, a_2, \dots, a_n\}$
 - ❑ Instancia (valores de los atributos): $X = \{x_1, x_2, \dots, x_K\}$
 - ❑ Probabilidades condicionales:
 - ❑ $P(c_j | X)$ Probabilidad de observar la clase c_j dada la instancia X
 - ❑ $P(X | c_j)$ Probabilidad de observar la instancia X dada la clase c_j

Clasificadores Bayesianos

□ Observaciones

Inst.	Trabajo	Ahorros	Vivienda	Duración	Clase
1	cualificado	> 20K	propia	24	bueno
2	cualificado	10 - 20K	alquilada	40	bueno
3	no cualificado	10 - 20K	propia	12	malo
4	autónomo	no conocidos	propia	36	bueno
5	no cualificado	0 - 10K	alquilada	12	bueno
6	autónomo	10 - 20K	alquilada	12	malo
7	no cualificado	0 - 10K	propia	30	malo
8	no cualificado	> 20K	alquilada	12	bueno

□ Problema de decisión

- $P(\text{clase} = \text{bueno} \mid \text{trabajo}=\text{cualificado}, \text{ahorros}=10 - 20\text{K}, \text{vivienda}=\text{propia}, \text{duración}=40)$
- $P(\text{clase} = \text{malo} \mid \text{trabajo}=\text{cualificado}, \text{ahorros}=10 - 20\text{K}, \text{vivienda}=\text{propia}, \text{duración}=40)$
- ¿bueno o malo?

El ejemplo de la lluvia y el paraguas

La predicción del tiempo para hoy es 20% de probabilidad de lluvia.

Estoy en una habitación sin ventanas y no puedo determinar si realmente llueve fuera. Sin embargo, veo a alguien que entra en la habitación y lleva un paraguas.

Sabiendo que la probabilidad de que alguien lleve paraguas es 70% si está lloviendo, y 10% si no está lloviendo, ¿cuál es la probabilidad de que realmente esté lloviendo? (Fuente: *inaugurelle rede*, Nijmegen Univ.; Prof. Heskes, 2009.)

Respuesta:

H = lluvia: “Está lloviendo”

D = paraguas: “Veo a alguien con un paraguas”

$P(H = \text{lluvia}) = 0.20$ **$P(H = \text{no lluvia}) = 0.80$** (prioris)

Verosimilitud

Probabilidades a posteriori:

$$P(D = \text{paraguas} \mid H = \text{no lluvia}) = 0.10$$

$$\mathbf{P(D = \text{paraguas} \mid H = \text{lluvia}) = 0.70}$$
 (solución ML)

Resultado final:

$$\begin{aligned} P(H = \text{llueve} \mid D = \text{paraguas}) &= \frac{P(D = \text{paraguas} \mid H = \text{llueve}) P(H = \text{llueve})}{P(D = \text{paraguas})} = \\ &= \frac{P(D = \text{paraguas} \mid H = \text{llueve}) P(H = \text{llueve})}{P(D = \text{paraguas} \mid H = \text{llueve}) P(H = \text{llueve}) + P(D = \text{paraguas} \mid H = \text{no llueve}) P(H = \text{no llueve})} = \frac{0.70 \times 0.20}{0.70 \times 0.20 + 0.10 \times 0.80} = 0.64 \end{aligned}$$

$$P(H = \text{no llueve} \mid D = \text{paraguas}) = 0.36$$

$$P(H = \text{no llueve} \mid D = \text{paraguas}) = 0.36$$

$$\mathbf{P(H = \text{llueve} \mid D = \text{paraguas}) = 0.64}$$
 (solución MAP)

El ejemplo del taxi

Ha habido un accidente de coche relacionado con un taxi y el taxista ha huído. Hay dos compañías de taxi en la ciudad: verdes (85%) y azules (15%).

¿Cuál es la probabilidad de que el taxi del accidente sea de la compañía azul?

(Fuente: Kahneman & Tversky, eds.: "Judgement under Uncertainty: Heuristics and biases", Cambridge Univ. Press, 1982).

Respuesta: $P(H = \text{azul}) = 0.15$ **$P(H = \text{verde}) = 0.85$** (prioris)

El ejemplo del taxi

¿Qué pasa si hay un testigo (80% de que diga la verdad) que dice que el taxi responsable del accidente era de la compañía azul?

H = azul: “El accidente fue causado por un taxi de la compañía azul”

D = azul: “El testigo dice que el taxi era azul”,

“15% de los taxis de la ciudad son azules” + “El grado de fiabilidad del testigo es del 80%”

$$P(D = azul \mid H = azul) = 0.80 \text{ (solución ML)}$$

Verosimilitud $P(D = verde \mid H = verde) = 0.80 \Rightarrow P(D = azul \mid H = verde) = 0.20$

Probabilidades

$$P(H = azul \mid D = azul) = \frac{P(D = azul \mid H = azul)P(H = azul)}{P(D = azul)} = \frac{0.80 \times 0.15}{P(D = azul)} = \frac{0.12}{P(D = azul)}$$
$$P(H = verde \mid D = azul) = \frac{P(D = azul \mid H = verde)P(H = verde)}{P(D = azul)} = \frac{0.20 \times 0.85}{P(D = azul)} = \frac{0.17}{P(D = azul)}$$

$$P(H = azul \mid D = azul) + P(H = verde \mid D = azul) = 1$$

Normalización: $\frac{0.12}{P(D = azul)} + \frac{0.17}{P(D = azul)} = 1 \rightarrow P(D = azul) = 0.29$

$$P(H = azul \mid D = azul) = 0.41$$

Resultado final: $P(H = verde \mid D = azul) = 0.59 \text{ (solución MAP)}$

El ejemplo de las lentes de contacto

❑ Atributos:

edad = {'joven', 'pre-présbita', 'présbita'}

prescripción = {'miope', 'hipermétrope'}

astigmatismo = {'sí', 'no'}

lagrimeo = {'reducido', 'normal'}

❑ Clase (¿se aconseja lentillas?): {'no' (n), 'blandas' (b), 'duras' (d)}

EDAD ('e')	PRESCRIPCIÓN ('p')	ASTIGMATISMO ('a')	LAGRIMEO ('l')	CLASE ('c')
joven	miope	no	reducido	no
joven	miope	no	normal	blandas
joven	miope	sí	reducido	no
joven	miope	sí	normal	duras
joven	hipermétrope	no	reducido	no
joven	hipermétrope	no	normal	blandas
joven	hipermétrope	sí	reducido	no
joven	hipermétrope	sí	normal	duras
pre-pres	miope	no	reducido	no
pre-pres	miope	no	normal	blandas
pre-pres	miope	sí	reducido	no
pre-pres	miope	sí	normal	duras
pre-pres	hipermétrope	no	reducido	no
pre-pres	hipermétrope	no	normal	blandas
pre-pres	hipermétrope	sí	reducido	no
pre-pres	hipermétrope	sí	normal	no
présbita	miope	no	reducido	no
présbita	miope	no	normal	no
présbita	miope	sí	reducido	no
présbita	miope	sí	normal	duras
présbita	hipermétrope	no	reducido	no
présbita	hipermétrope	no	normal	blandas
présbita	hipermétrope	sí	reducido	no
présbita	hipermétrope	sí	normal	no

Probabilidades

□ **Ejemplo 1: se observa prescripción='miope', lagrimeo = 'normal'**

□ **Probabilidades a priori de la clase**

$$P(c = 'n') = \frac{15}{24} \quad P(c = 'b') = \frac{5}{24} \quad P(c = 'd') = \frac{4}{24}$$

□ **Verosimilitud** (probabilidad de una observación condicionada a la clase)
prescripción='miope', lagrimeo='normal'

$$P(p = 'm', l = 'n' | c = 'n') = \frac{1}{15}$$

$$P(p = 'm', l = 'n' | c = 'b') = \frac{2}{5}$$

$$(p = 'm', l = 'n' | c = 'd') = \frac{3}{4} \quad [\text{ML}]$$

□ **Probabilidades a posteriori de la clase (condicionadas en las observaciones)**

$$P(c = 'n' | p = 'm', l = 'n') = \frac{P(p = 'm', l = 'n' | c = 'n')P(c = 'n')}{P(p = 'm', l = 'n')} = \frac{\frac{1}{15} \times \frac{15}{24}}{P(p = 'm', l = 'n')} = \frac{1}{6} [17\%]$$

$$P(c = 'b' | p = 'm', l = 'n') = \frac{P(p = 'm', l = 'n' | c = 'b')P(c = 'b')}{P(p = 'm', l = 'n')} = \frac{\frac{2}{5} \times \frac{5}{24}}{P(p = 'm', l = 'n')} = \frac{1}{3} [33\%]$$

$$P(c = 'd' | p = 'm', l = 'n') = \frac{P(p = 'm', l = 'n' | c = 'd')P(c = 'd')}{P(p = 'm', l = 'n')} = \frac{\frac{3}{4} \times \frac{4}{24}}{P(p = 'm', l = 'n')} = \frac{1}{2} [50\%]$$

$$P(p = 'm', l = 'n') = \frac{1}{24} + \frac{2}{24} + \frac{3}{24} = \frac{1}{4}$$

Probabilidades

❑ **Ejemplo 2: se observa prescripción='miope', lagrimeo = 'reducido'**

❑ **Probabilidades a priori de la clase**

$$P(c = 'n') = \frac{15}{24} \quad P(c = 'b') = \frac{5}{24} \quad P(c = 'd') = \frac{4}{24}$$

❑ **Verosimilitud** (probabilidad de una observación condicionada a la clase)
prescripción='miope', lagrimeo='reducido'

$$P(p = 'm', l = 'r' | c = 'n') = \frac{6}{15} \quad [\text{ML}]$$

$$P(p = 'm', l = 'r' | c = 'b') = \frac{0}{5}$$

$$P(p = 'm', l = 'r' | c = 'd') = \frac{0}{4}$$

❑ **Probabilidades a posteriori de la clase (condicionadas en las observaciones)**

$$P(c = 'n' | p = 'm', l = 'r') = \frac{P(p = 'm', l = 'r' | c = 'n')P(c = 'n')}{P(p = 'm', l = 'r')} = \frac{\frac{6}{15} \times \frac{15}{24}}{\frac{6}{24}} = 1[100\%]$$

$$P(c = 'b' | p = 'm', l = 'r') = \frac{P(p = 'm', l = 'r' | c = 'b')P(c = 'b')}{P(p = 'm', l = 'r')} = \frac{\frac{0}{5} \times \frac{5}{24}}{\frac{6}{24}} = 0[0\%]$$

$$P(c = 'd' | p = 'm', l = 'r') = \frac{P(p = 'm', l = 'r' | c = 'd')P(c = 'd')}{P(p = 'm', l = 'r')} = \frac{\frac{0}{4} \times \frac{4}{24}}{\frac{6}{24}} = 0[0\%]$$

$$P(p = 'm', l = 'r') = \frac{6}{24} + \frac{0}{24} + \frac{0}{24} = \frac{1}{4}$$

Probabilidades

❑ **Ejemplo 3: se observa prescripción='hipermétrope', lagrimeo = 'normal'**

❑ **Probabilidades a priori de la clase**

$$P(c = 'n') = \frac{15}{24} \quad P(c = 'b') = \frac{5}{24} \quad P(c = 'd') = \frac{4}{24}$$

❑ **Verosimilitud** (probabilidad de una observación condicionada a la clase)
prescripción='hipermétrope', lagrimeo='normal'

$$P(p = 'h', l = 'n' | c = 'n') = \frac{2}{15}$$

$$P(p = 'h', l = 'n' | c = 'b') = \frac{3}{5} \quad [\text{ML}]$$

$$P(p = 'h', l = 'n' | c = 'd') = \frac{1}{4}$$

❑ **Probabilidades a posteriori de la clase (condicionadas en las observaciones)**

$$P(c = 'n' | p = 'h', l = 'n') = \frac{P(p = 'h', l = 'n' | c = 'n')P(c = 'n')}{P(p = 'm', l = 'n')} = \frac{\frac{2}{15} \times \frac{15}{24}}{P(p = 'h', l = 'n')} = \frac{1}{3} [33\%]$$

$$P(c = 'b' | p = 'h', l = 'n') = \frac{P(p = 'h', l = 'n' | c = 's')P(c = 's')}{P(p = 'm', l = 'n')} = \frac{\frac{3}{5} \times \frac{5}{24}}{P(p = 'h', l = 'n')} = \frac{1}{2} [50\%]$$

$$P(c = 'd' | p = 'h', l = 'n') = \frac{P(p = 'h', l = 'n' | c = 'h')P(c = 'h')}{P(p = 'h', l = 'n')} = \frac{\frac{1}{4} \times \frac{4}{24}}{P(p = 'h', l = 'n')} = \frac{1}{6} [17\%]$$

$$P(p = 'h', l = 'n') = \frac{2}{24} + \frac{3}{24} + \frac{1}{24} = \frac{1}{4}$$

Probabilidades

❑ **Ejemplo 4:** se observa prescripción='hipermétrope', lagrimeo = 'reducido'

❑ **Probabilidades a priori de la clase**

$$P(c = 'n') = \frac{15}{24} \quad P(c = 'b') = \frac{5}{24} \quad P(c = 'd') = \frac{4}{24}$$

❑ **Verosimilitud** (probabilidad de una observación condicionada a la clase)
prescripción='hipermétrope', lagrimeo='reducido'

$$P(p = 'h', l = 'r' | c = 'n') = \frac{6}{15} \quad [\text{ML}]$$

$$P(p = 'h', l = 'r' | c = 'b') = \frac{0}{5}$$

$$P(p = 'h', l = 'r' | c = 'd') = \frac{0}{4}$$

❑ **Probabilidades a posteriori de la clase (condicionadas en las observaciones)**

$$P(c = 'n' | p = 'h', l = 'r') = \frac{P(p = 'h', l = 'r' | c = 'n')P(c = 'n')}{P(p = 'h', l = 'r')} = \frac{\frac{6}{15} \times \frac{15}{24}}{\frac{6}{15} \times \frac{15}{24}} = 1[100\%]$$

$$P(c = 'b' | p = 'h', l = 'r') = \frac{P(p = 'h', l = 'r' | c = 'b')P(c = 'b')}{P(p = 'h', l = 'r')} = \frac{\frac{0}{5} \times \frac{5}{24}}{\frac{6}{15} \times \frac{15}{24}} = 0[0\%]$$

$$P(c = 'd' | p = 'h', l = 'r') = \frac{P(p = 'h', l = 'r' | c = 'd')P(c = 'd')}{P(p = 'h', l = 'r')} = \frac{\frac{0}{4} \times \frac{4}{24}}{\frac{6}{15} \times \frac{15}{24}} = 0[0\%]$$

$$P(p = 'h', l = 'r') = \frac{6}{24} + \frac{0}{24} + \frac{0}{24} = \frac{1}{4}$$

Clasificador ML vs. Clasificador Bayes

- ❑ **Clasificador de Máxima Verosimilitud:** Asigna la clase que maximiza la verosimilitud (probabilidad de la observación condicionada a la clase) **[ML]**
- ❑ **Clasificador de Bayes:** Asigna la clase cuya probabilidad a posteriori (dada la observación) es máxima **[MAP]**

prescripción	lagrimeo	clase predicha (ML)	clase predicha (Bayes)
<i>miope</i>	<i>normal</i>	<i>duras</i>	<i>duras</i> [50%]
<i>miope</i>	<i>reducido</i>	<i>no</i>	<i>no</i> [100%]
<i>hipermétrope</i>	<i>normal</i>	<i>blandas</i>	<i>blandas</i> [50%]
<i>hipermétrope</i>	<i>reducido</i>	<i>no</i>	<i>no</i> [100%]

- ❑ **Prioris uniformes \Rightarrow Clasificador ML = Clasificador Bayes**
- ❑ **En general, las predicciones del clasificador ML pueden ser diferentes que las del clasificador de Bayes.**
- ❑ **Bayes es óptimo (minimiza el error).**

Clasificadores Bayesianos: Naïve Bayes

- Recordamos el Teorema de Bayes:

$$P(c_j | x_j) = \frac{P(x_j | c_j)P(c_j)}{P(x_j)}$$

- La probabilidad a priori de una instancia x_i es independiente del valor de la clase, por lo que generalmente $P(x_i)$ no se calcula → **Naïve**
- La idea del clasificador consiste en elegir la clase más probable según la probabilidad a posteriori $P(c_j | x_i) = P(H | D)$

$$\text{ClasificadorBayesiano}(x_i) = \operatorname{argmax}_{c_j \in C} P(x_i | c_j)P(c_j) = \operatorname{argmax}_{c_j \in C} P(c_j) \prod_{i \in A}^n P(x_i | c_j)$$

Ejemplo: Clasificador Bayesiano

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
Day1	Sunny	Hot	High	Weak	No
Day2	Sunny	Hot	High	Strong	No
Day3	Overcast	Hot	High	Weak	Yes
Day4	Rain	Mild	High	Weak	Yes
Day5	Rain	Cool	Normal	Weak	Yes
Day6	Rain	Cool	Normal	Strong	No
Day7	Overcast	Cool	Normal	Strong	Yes
Day8	Sunny	Mild	High	Weak	No
Day9	Sunny	Cool	Normal	Weak	Yes
Day10	Rain	Mild	Normal	Weak	Yes
Day11	Sunny	Mild	Normal	Strong	Yes
Day12	Overcast	Mild	High	Strong	Yes
Day13	Overcast	Hot	Normal	Weak	Yes
Day14	Rain	Mild	High	Strong	No

Ejemplo: Clasificador Naïve Bayes

$x = \langle \text{Outlook}=\text{Sunny}, \text{Temp}=\text{Cool}, \text{Hum}=\text{High}, \text{Wind}=\text{Strong} \rangle$

$h_{NB} = \operatorname{argmax} P(h) P(x|h)$

$= \operatorname{argmax} P(h) \prod P(V_i | h)$

$= \operatorname{argmax} P(h) P(\text{Outlook}=\text{Sunny} | h) P(\text{Temp}=\text{Cool} | h) \\ P(\text{Hum}=\text{High} | h) P(\text{Wind}=\text{Strong} | h)$

Aproximando las probabilidades por la frecuencia:

$P(\text{PlayTennis} = \text{yes}) = 9/14 = 0.64$

$P(\text{PlayTennis} = \text{no}) = 5/14 = 0.36$

$P(\text{Wind} = \text{Strong} | \text{PlayTennis} = \text{yes}) = 3/9 = 0.33$

$P(\text{Wind} = \text{Strong} | \text{PlayTennis} = \text{no}) = 3/5 = 0.60$

Aplicandolo a las fórmulas:

$P(\text{yes}) P(\text{Sunny}|\text{yes}) P(\text{Cool}|\text{yes}) P(\text{High}|\text{yes}) P(\text{String}|\text{yes}) = 0.0053$

$P(\text{no}) P(\text{Sunny}|\text{no}) P(\text{Cool}|\text{no}) P(\text{High}|\text{no}) P(\text{String}|\text{no}) = \mathbf{0.0206}$

\Rightarrow Answer: **PlayTennis = no**

\Rightarrow **Con 79.5% de certeza**

Clasificador Naïve Bayes

❑ **Ejemplo 1 del problema de las lentillas: prescr. = 'miope', lagr. = 'normal'**

❑ **Prioris de la clase** $P(c = 'n') = \frac{15}{24}$ $P(c = 'b') = \frac{5}{24}$ $P(c = 'd') = \frac{4}{24}$

❑ **Marginales condicionados:**

	no	blandas	duras
p = 'miope'	7	2	3
p = 'hipermétrope'	8	3	1

$$P(p = 'm' | c = 'n') = \frac{7}{15}$$

$$P(p = 'm' | c = 'b') = \frac{2}{5}$$

$$P(p = 'm' | c = 'd') = \frac{3}{4}$$

	no	blandas	duras
l = 'normal'	3	5	4
l = 'reducido'	12	0	0

$$P(l = 'n' | c = 'n') = \frac{3}{15}$$

$$P(l = 'n' | c = 'b') = \frac{5}{5}$$

$$P(l = 'n' | c = 'd') = \frac{4}{4}$$

❑ **Suposición de**

independencia condicional:

$$P(p = 'm', l = 'n' | c = 'n') \approx P(p = 'm' | c = 'n') \cdot P(l = 'n' | c = 'n') = \frac{7}{15} \times \frac{3}{15}$$

$$P(p = 'm', l = 'n' | c = 'b') \approx P(p = 'm' | c = 'b') \cdot P(l = 'n' | c = 'b') = \frac{2}{5} \times \frac{5}{5}$$

$$P(p = 'm', l = 'n' | c = 'd') \approx P(p = 'm' | c = 'd') \cdot P(l = 'n' | c = 'd') = \frac{3}{4} \times \frac{4}{4}$$

❑ **Naïve Bayes:**

$$P(c = 'n' | p = 'm', l = 'n') = \frac{P(p = 'm', l = 'n' | c = 'n') P(c = 'n')}{P(p = 'm', l = 'n')} \approx \frac{\frac{7}{15} \times \frac{3}{15} \times \frac{15}{24}}{\text{Norm}} = 0.22 \quad [\text{exacta: } 17\%]$$

$$P(c = 'b' | p = 'm', l = 'n') = \frac{P(p = 'm', l = 'n' | c = 'b') P(c = 'b')}{P(p = 'm', l = 'n')} \approx \frac{\frac{2}{5} \times \frac{5}{5} \times \frac{5}{24}}{\text{Norm}} = 0.31 \quad [\text{exacta: } 33\%]$$

$$P(c = 'd' | p = 'm', l = 'n') = \frac{P(p = 'm', l = 'n' | c = 'd') P(c = 'd')}{P(p = 'm', l = 'n')} \approx \frac{\frac{3}{4} \times \frac{4}{4} \times \frac{4}{24}}{\text{Norm}} = 0.47 \quad [\text{exacta: } 50\%]$$

$$P(p = 'm', l = 'n') \approx \text{Norm} = \frac{7}{120} + \frac{1}{12} + \frac{1}{8} = \frac{4}{15}$$

Naïve Bayes

❑ Ventajas

- ❑ A pesar del fuerte supuesto de independencia condicional, funciona sorprendentemente bien en muchos problemas del mundo real.
- ❑ Incluso si existen dependencias, sus efectos pueden anularse.
- ❑ Entrenamiento y predicción rápidos.

❑ Inconvenientes

- ❑ Las estimaciones de probabilidad no son confiables

https://scikit-learn.org/stable/modules/naive_bayes.html

Estimación de probabilidades

- ❑ Las estimaciones de frecuencia de probabilidades pueden no ser fiables, especialmente cuando las muestras son pequeñas.

	c = "no"	c = "soft"	c = "hard"
tr = "normal"	3	5	4
tr = "reduced"	12	0	0

$$P(tr = "r" | c = "n") = \frac{12}{15}$$

$$P(tr = "r" | c = "s") = \frac{0}{5}$$

$$P(tr = "r" | c = "h") = \frac{0}{4}$$

Esta probabilidad es cero. Por lo tanto, todos los productos que involucrar este término (por ejemplo, en NB) será cero, que no es razonable.

Estimador de Laplace

Estimación de probabilidades con frecuencias:

$$P_i = \frac{n_i}{n_{total}}, \quad i = 1, 2, \dots, K$$

Estimador de Laplace: Añadir ejemplos ficticios

$$P_i = \frac{n_i + \mu / K}{n_{total} + \mu}, \quad i = 1, 2, \dots, K$$

Evita estimaciones nulas para las probabilidades

Estimaciones más robustos

Asintóticamente pequeño

	no	blandas	duras
l = 'normal'	3+1	5+1	4+1
l = 'reducido'	12+1	0+1	0+1

$$P(l = 'r' | c = 'n') = \frac{12+1}{15+2} = \frac{13}{17}$$

$$P(l = 'r' | c = 'b') = \frac{0+1}{5+2} = \frac{1}{7}$$

$$P(l = 'r' | c = 'd') = \frac{0+1}{4+2} = \frac{1}{6}$$

Hasta ahora sabemos

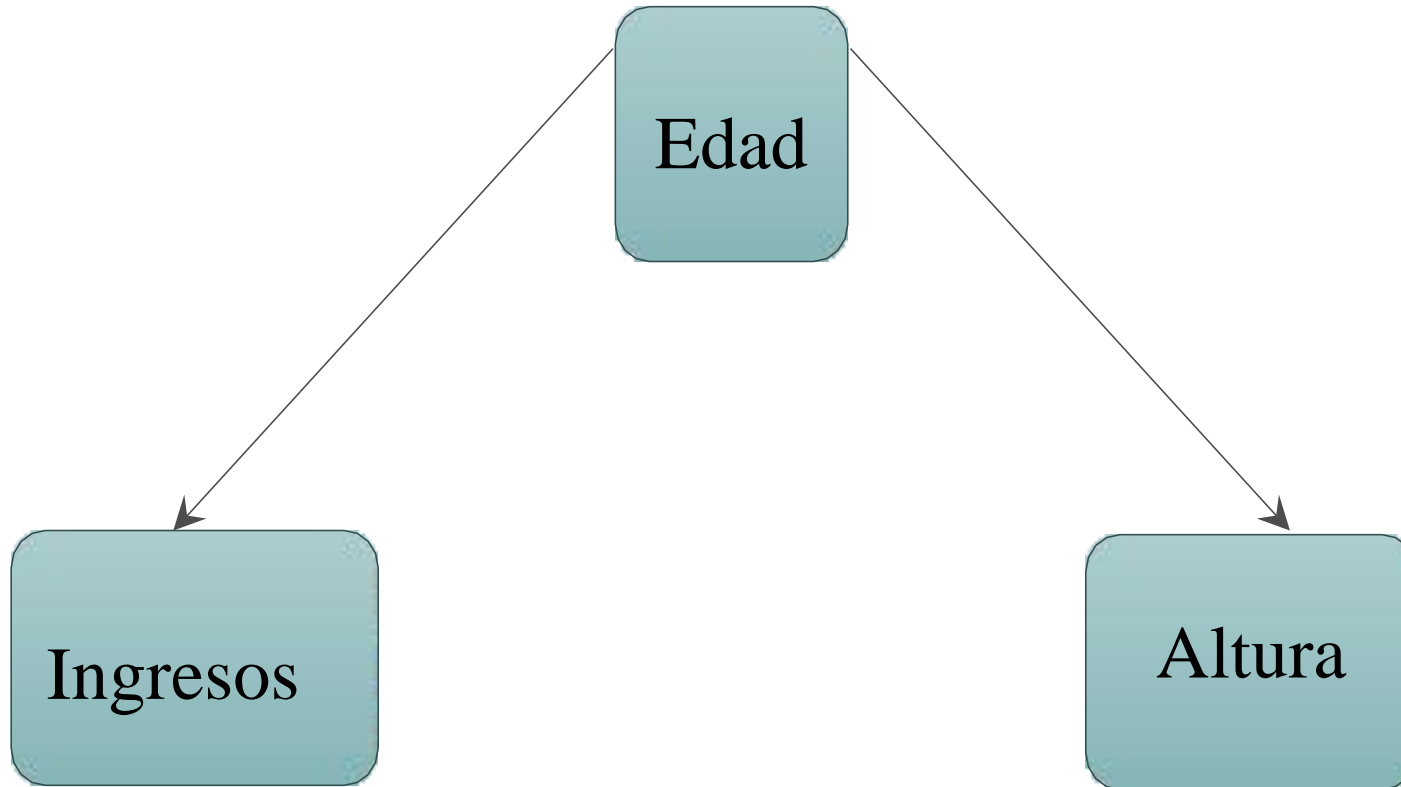
- ❑ Representación en dominios con variables aleatorias + distribución de probabilidad
 - ❑ Dada la distribución de probabilidad para todos los eventos posibles, podemos resolver consultas $P(\text{Variables} \mid \text{Observación})$
 - ❑ El tamaño de distribución es exponencial en el número de variables.
 - ❑ La independencia podría permitirnos un razonamiento más eficiente
-
- ❑ Hoy: ¿cómo podemos usar las probabilidades de manera más eficiente?
 - ❑ Respuesta: redes bayesianas

Incertidumbre en IA

- ❑ Formalización de incertidumbre mediante probabilidades
- ❑ Teorema de Bayes en IA
- ❑ **Redes Bayesianas**

Independencia condicional

□ $P(\text{Ingresos} \mid \text{Altura, Edad}) = P(\text{Ingresos} \mid \text{Edad})$



□ Ingresos y altura son condicionalmente independientes "dada" Edad

Independencia condicional

□ $P(\text{Calzado} \mid \text{Altura, Edad}) = P(\text{Calzado} \mid \text{Altura})$



□ Calzado y edad son condicionalmente independientes "dada" Altura

Independencia condicional

- ❑ X e Y son condicionalmente independientes dado Z si
 - ❑ $P(X, Y | Z) = P(X | Z) P(Y | Z)$
- ❑ También:
 - ❑ $P(X | Y, Z) = P(X | Z)$
- ❑ A menudo reduce el número de parámetros de exponencial en n (número de variables) a lineal en n
- ❑ La independencia condicional es la herramienta de razonamiento probabilístico eficiente
 - ❑ menos parámetros
 - ❑ menos cómputo
- ❑ Está representado por los ejes que faltan

4.3 Definición de una red Bayesiana

- ❑ Un conjunto de **nodos**
 - ❑ cada nodo representa una variable aleatoria
 - ❑ las variables pueden ser discretas o continuas
- ❑ Un conjunto de **ejes**
 - ❑ un borde del nodo X al nodo Y: X tiene una influencia directa en Y
 - ❑ es un grafo acíclico directo (DAG)
- ❑ **Distribuciones de probabilidad**
 - ❑ cada nodo X tiene una tabla de probabilidad condicional (CPT) que define los efectos de sus padres

$$P(\text{Nodo} \mid \text{Padres}(\text{Nodo}))$$

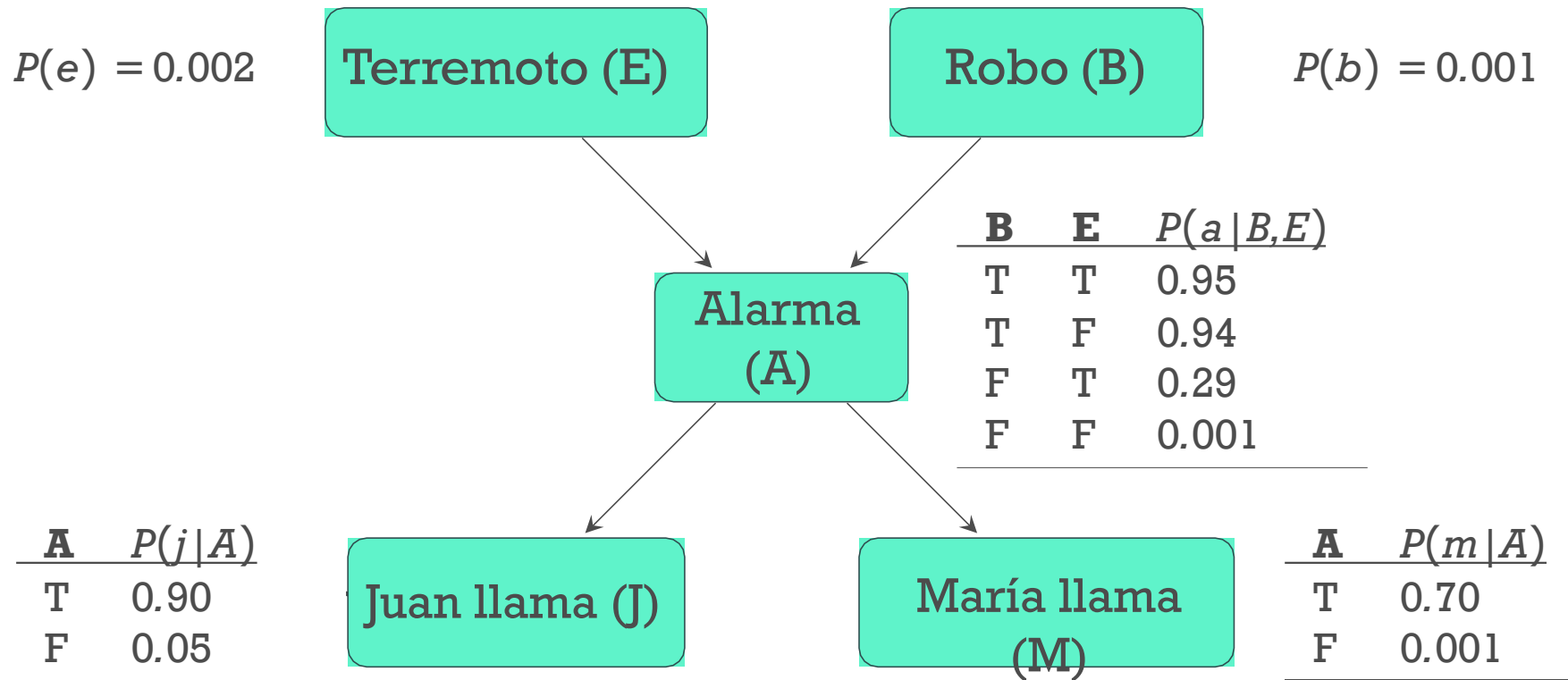
- ❑ los padres del nodo X son los únicos ejes dirigidos a X
- ❑ si un nodo no tiene padres, es la probabilidad "a priori"

$$P(\text{nodo})$$

Ejemplo de una alarma

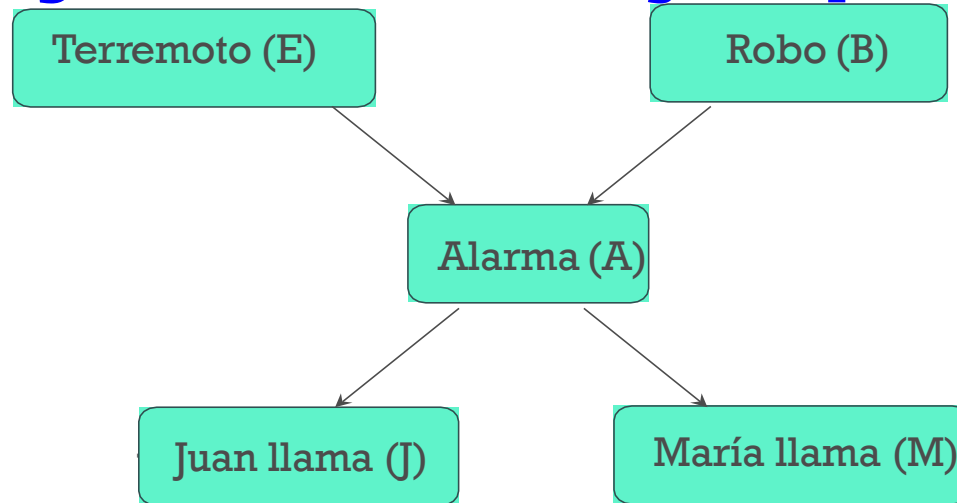
- ❑ Tenemos un sistema antirrobo en casa con alarma.
- ❑ Detecta ladrones, pero la alarma también se dispara con algunos terremotos.
- ❑ Hay dos vecinos (Juan y María) que nos llamarán si escuchan la alarma.
- ❑ Juan siempre llama cuando escucha la alarma, pero a veces se confunde con un timbre
- ❑ María escucha música muy fuerte, así que a veces no puede escuchar la alarma.
- ❑ Modelando:
 - ❑ Terremoto: E . $E = \text{Verdadero} \sim e$. $E = \text{Falso} \sim \neg e$
 - ❑ Robo: B . $B = \text{Verdadero} \sim b$. $B = \text{Falso} \sim \neg b$
 - ❑ Alarma: A ($a, \neg a$)
 - ❑ Juan llama: J ($j, \neg j$)
 - ❑ María llama: M ($m, \neg m$)

Red Bayesiana del ejemplo



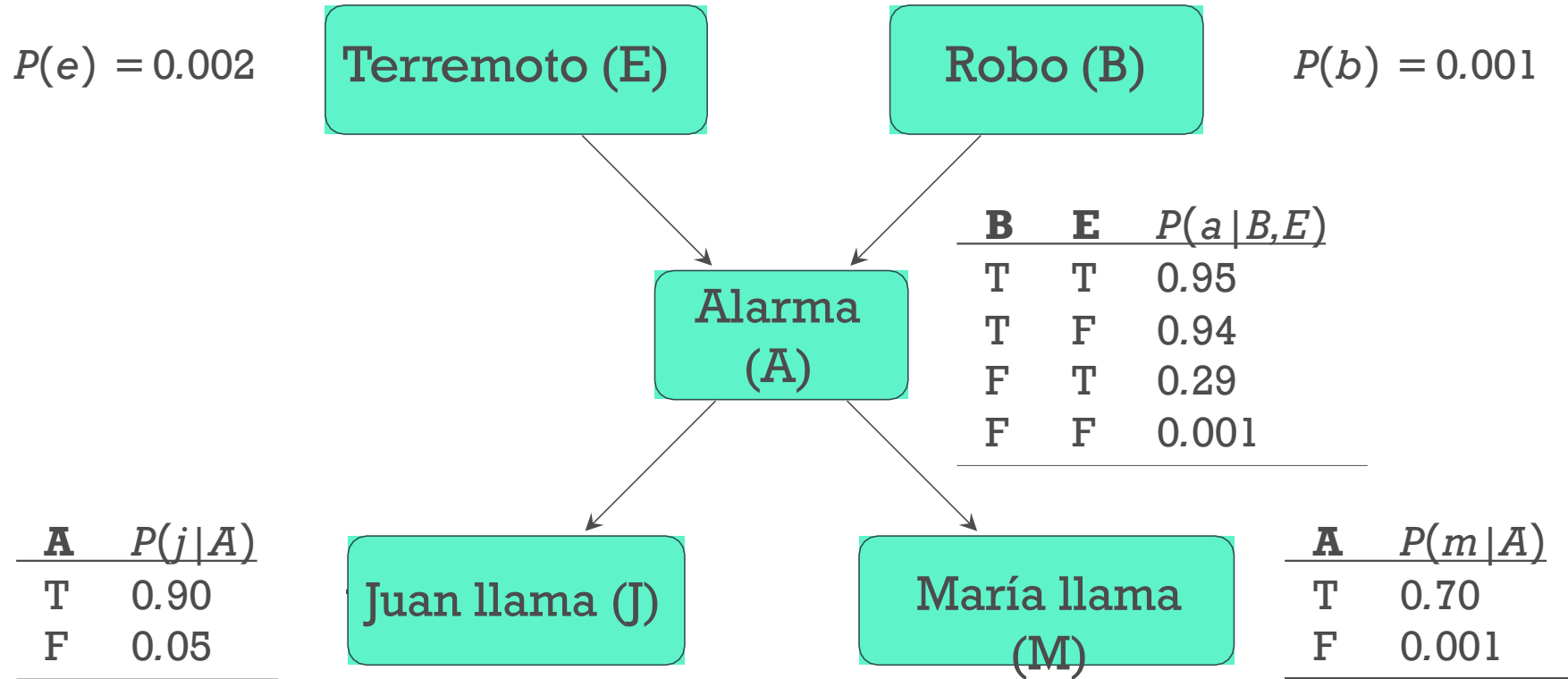
- ❑ Solo proporcionamos $P(e)$ dado que $P(\neg e) = 1 - P(e)$
- ❑ Además, $P(\neg a | b, \neg e) = 1 - P(a | b, \neg e)$
- ❑ La red refleja las causas directas de sus variables:
 - ❑ un ladrón puede disparar la alarma
 - ❑ un terremoto puede disparar la alarma
 - ❑ la alarma puede hacer que Maria llame
 - ❑ la alarma puede hacer que Juan llame

Red Bayesiana del ejemplo



- ❑ No hay dependencia entre terremoto y robo
- ❑ Pero, hay dependencia entre Alarma y las otras dos variables:
 - ❑ $P(\text{Alarma} \mid \text{Terremoto}, \text{Robo}) = P(\text{Alarma} \mid \text{Terremoto})$
 - ❑ $P(\text{Alarma} \mid \text{Terremoto}, \text{Robo}) = P(\text{Alarma} \mid \text{Robo})$
- ❑ Existe una independencia condicional entre las llamadas de Juan y María y las variables Terremoto y Robo, dada la variable Alarma
 - ❑ $P(\text{Juan} \mid \text{Alarma}, \text{Terremoto}, \text{Robo}) = P(\text{Juan} \mid \text{Alarma})$
 - ❑ $P(\text{Maria} \mid \text{Alarm}, \text{Earthquake}, \text{Robbery}) = P(\text{Maria} \mid \text{Alarm})$

Red Bayesiana: compacta



- ❑ La distribución conjunta explícita requeriría $2^5 - 1 = 31$ parámetros
- ❑ La red bayesiana utiliza $1 + 1 + 4 + 2 + 2 = 10$ parámetros

Semántica de las redes bayesianas

- Semántica global: la distribución de probabilidad conjunta es el producto de distribuciones locales.

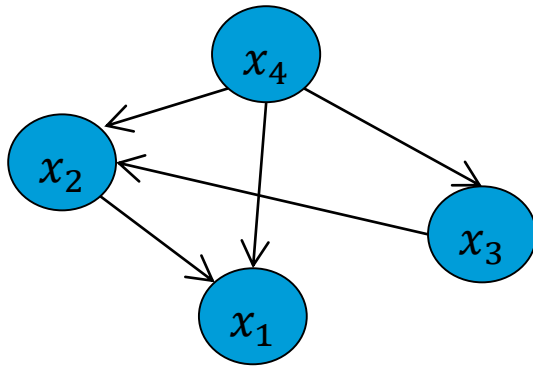
$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1}) = \prod_{i=1}^n P(X_i | \text{Padres}(X_i))$$

- Origen: Regla de la cadena

$$\begin{aligned} P(X_1, X_2, X_3, X_4) &= P(X_1 | X_2, X_3, X_4) P(X_2, X_3, X_4) = \\ &P(X_1 | X_2, X_3, X_4) P(X_2 | X_3, X_4) P(X_3, X_4) = \\ &P(X_1 | X_2, X_3, X_4) P(X_2 | X_3, X_4) P(X_3 | X_4) P(X_4) \end{aligned}$$

Interpretación del grafo

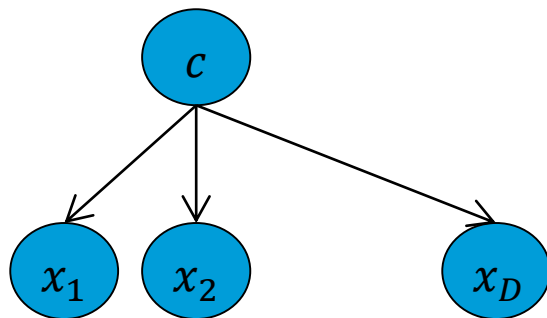
$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Padres}(X_i))$$



$$\square P(x_1, x_2, x_3, x_4) = P(x_1 \mid x_2, x_4) P(x_2 \mid x_3, x_4) P(x_3 \mid x_4) P(x_4)$$

Naïve Bayes

$$P(\mathbf{x}, c) = \prod_{i=1}^N P(x_i|c)P(c)$$



Problema de las lentes de contacto

❑ **Ejemplo:** se observa edad='joven', astigmatismo='sí', lagrimeo = 'normal'

❑ **Prioris de las clases** $P(c = 'n') = \frac{15}{24}$ $P(c = 'b') = \frac{5}{24}$ $P(c = 'd') = \frac{4}{24}$

❑ **Marginales condicionados:**

	no	blandas	duras
e = 'j'	4	2	2
Total	15	5	4

$$P(e = 'm' | c = 'n') = \frac{4}{15}$$

$$P(e = 'm' | c = 'b') = \frac{2}{5}$$

$$P(e = 'm' | c = 'd') = \frac{2}{4}$$

	no	blandas	duras
as = 'sí'	8	0	4
Total	15	5	4

$$P(as = 'sí' | c = 'n') = \frac{4}{15}$$

$$P(as = 'sí' | c = 'b') = \frac{2}{5}$$

$$P(as = 'sí' | c = 'd') = \frac{2}{4}$$

	no	blandas	duras
l = 'normal'	3	5	4
Total	15	5	4

$$P(l = 'n' | c = 'n') = \frac{3}{15}$$

$$P(l = 'n' | c = 'b') = \frac{5}{5}$$

$$P(l = 'n' | c = 'd') = \frac{4}{4}$$

❑ **Naïve Bayes. Asumimos independencia**

$$P(c = 'n' | e = 'j', as = 'y', t = 'n') = \frac{P(e = 'j', as = 'sí', l = 'n' | c = 'n')P(c = 'n')}{P(e = 'j', as = 'sí', l = 'n')} \approx \frac{\frac{4}{15} \times \frac{8}{15} \times \frac{3}{15} \times \frac{15}{24}}{Norm} = 0.1758 [18\%]$$

$$P(c = 'b' | e = 'j', as = 'sí', l = 'n') = \frac{P(e = 'j', as = 'sí', l = 'n' | c = 'b')P(c = 'b')}{P(e = 'j', as = 'sí', l = 'n')} \approx \frac{\frac{2}{5} \times \frac{0}{5} \times \frac{5}{5} \times \frac{5}{24}}{Norm} = 0.00 [0\%]$$

$$P(c = 'd' | e = 'j', as = 'sí', l = 'n') = \frac{P(e = 'j', as = 'sí', l = 'n' | c = 'd')P(c = 'd')}{P(e = 'j', as = 'sí', l = 'n')} \approx \frac{\frac{2}{4} \times \frac{4}{4} \times \frac{4}{4} \times \frac{4}{24}}{Norm} = 0.8242 [82\%]$$

$$P(e = 'j', as = 'sí', l = 'n') \approx Norm = 0.1011$$

Un modelo más complejo

⌘ **Ejemplo:** se observa edad='joven', astigmatismo='sí', lagrimeo = 'normal'

⌘ **Prioris de las clases** $P(c = 'n') = \frac{15}{24}$ $P(c = 'b') = \frac{5}{24}$ $P(c = 'd') = \frac{4}{24}$

⌘ **Modelo:** $P(e, as, l | c) = P(e | c) \cdot P(as | e, c) \cdot P(l | c)$

⌘ **Marginales condicionados:**

	no	blandas	duras
l='normal',	3	5	4
Total	15	5	4

$$P(l = 'n' | c = 'n') = \frac{3}{15}$$

$$P(l = 'n' | c = 'b') = \frac{5}{5}$$

$$P(l = 'n' | c = 'd') = \frac{4}{4}$$

	c = 'no' e = 'joven'	c = 'blandas' e = 'joven'	c = 'duras' e = 'joven'
as='sí'	2	0	2
Total	4	2	2

$$P(as = 'sí' | e = 'j', c = 'n') = \frac{2}{4} = \frac{1}{2}$$

$$P(as = 'sí' | e = 'j', c = 'b') = \frac{0}{2}$$

$$P(as = 'sí' | e = 'j', c = 'd') = \frac{2}{2}$$

	no	blandas	duras
e = 'joven'	4	2	2
Total	15	5	4

$$P(e = 'j' | c = 'n') = \frac{4}{15}$$

$$P(e = 'j' | c = 'b') = \frac{2}{5}$$

$$P(e = 'j' | c = 'd') = \frac{2}{4}$$

$$P(c = 'n' | e = 'j', as = 'sí', l = 'n') = \frac{P(e = 'j', as = 'sí', l = 'n' | c = 'n') P(c = 'n')}{P(e = 'j', as = 'sí', l = 'n')} \approx \frac{\frac{4}{15} \times \frac{2}{4} \times \frac{3}{15} \times \frac{15}{24}}{Norm} = \frac{1}{6} \quad [17\%]$$

$$P(c = 'b' | e = 'j', as = 'sí', l = 'n') = \frac{P(e = 'j', as = 'sí', l = 'n' | c = 'b') P(c = 'b')}{P(e = 'j', as = 'sí', l = 'n')} \approx \frac{\frac{2}{5} \times \frac{0}{2} \times \frac{5}{5} \times \frac{5}{24}}{Norm} = 0 \quad [0\%]$$

$$P(c = 'd' | e = 'j', as = 'sí', l = 'n') = \frac{P(e = 'j', as = 'sí', l = 'n' | c = 'd') P(c = 'd')}{P(e = 'j', as = 'sí', l = 'n')} \approx \frac{\frac{2}{4} \times \frac{2}{2} \times \frac{4}{4} \times \frac{4}{24}}{Norm} = \frac{5}{6} \quad [83\%]$$

$$P(e = 'j', as = 'sí', l = 'n') \approx Norm = \frac{1}{60} + 0 + \frac{1}{12} = \frac{1}{10}$$

