Tema 4
Incertidumbre en IA

4.1 Formalización de incertidumbre mediante probabilidades

4.2 Teorema de Bayes en IA

4.3 Redes Bayesianas

Lecturas recomendadas:
- CAPÍTULO 1 del libro de Bishop
- CAPÍTULOS 1, 2 del libro de Jaynes

4.1 Formalización de incertidumbre mediante probabilidades: Probabilidades y razonamiento

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 y plausibilidad

P(A | I)

Grado de plausibilidad de una aserción A, dada la información I

1. $P(Verdad \mid I) = 1, \qquad P(Falso \mid I) = 0$
2. $0 \leq P(A \mid I) \leq 1$
3. Regla de la suma: $P(A \mid I) + P(\overline{A} \mid I) = 1$

4. Regla del producto: $P(A,B \mid I) = P(B \mid A,I) \cdot P(A \mid I)$

5. Teorema de Bayes: $$P(B \mid A,I) = \frac{P(A \mid B,I)P(B \mid I)}{P(A \mid I)}$$

Dem. $$\left.\begin{array}{l} P(A,B \mid I) = P(A \mid B,I)P(B \mid I) \\ P(A,B \mid I) = P(B \mid A,I)P(A \mid I) \end{array}\right\} \to P(A \mid B,I)P(B \mid I) = P(B \mid A,I)P(A \mid I)$$

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

[NOTA: en notación de lógica proposicional

P(A | I) + P(¬A | I) = 1

P(A ∧ B | I) = P(A | B ∧ I) P(B | I)]

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. $\neg$BATERÍA_OK $\Rightarrow$ $\neg$FUNCIONA (causal?)
  "El dispositivo no funciona porque la batería no está OK"
  FUNCIONA $\Rightarrow$ 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, *Nubes* $\Rightarrow$ *Lluvia* es incorrecto.
  La aserción correcta es *Lluvia* $\Rightarrow$ *Nubes,* que no puede ser entendido 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 de la urna y no se devuelve a ella.
    En una segunda extracción se extrae una bola roja.
    Dado que sólo hay una bola roja en la urna, la probabilidad de haber observado una bola roja en la primera extracción es 0.

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

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

4.2 Teorema de Bayes en IA

$$P(H \mid D,I) = \frac{P(D \mid H,I)P(H \mid I)}{P(D \mid I)}$$

H : Hipótesis

D : Datos

I : Información adicional

$$P(D \mid I) = \int dH \; P(D \mid H,I)P(H \mid I)$$

P(H | I): **Probabilidad a priori** de la hipótesis.

Probabilidad de la hipótesis, antes de observar los datos.

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

P(D | I): **Evidencia** de los datos.

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

P(H | D, I): **Probabilidad a posteriori** de la hipótesis.

Probabilidad de la hipótesis, después de observar los datos.

Inferencia a partir de los datos

- Máxima verosimilitud (ML):
  Selecciona la hipótesis que maximiza la verosimilitud de la hipótesis dados los datos

$$H^* = \arg\max_H P(D \mid H, I)$$

  - ML no usa información de los prioris (equivale a asumir un priori uniforme, *P(H|I) = constante*).
- Máxima probabilidad a posteriori (MAP)
  Selecciona la hipótesis que maximiza la probabilidad a posteriori

$$H^* = \arg\max_H P(H \mid D, I) = \arg\max_H \frac{P(D \mid H, I) P(H \mid I)}{P(D \mid I)} =$$
$$= \arg\max_H P(D \mid H, I) P(H \mid I)$$

- Inferencia Bayesiana.
  Promediar sobre todas las hipótesis con probabilidades

$$P(H \mid D, I) = \frac{P(D \mid H, I) P(H \mid I)}{P(D \mid I)}$$
$$E[f(H)] = \int dH \; f(H) P(H \mid D, I)$$

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 = lluvia | I) = 0.20 **P(H = no lluvia | I) = 0.80** (prioris)

Verosimilitud

$$P(D = paraguas \mid H = no\ lluvia, I) = 0.10$$

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

Probabilidades a posteriori:

$$P(H = llueve \mid D = paraguas, I) = \frac{P(D = paraguas \mid H = llueve, I) P(H = llueve \mid I)}{P(D = paraguas \mid I)} =$$

$$= \frac{P(D = paraguas \mid H = llueve, I) P(H = llueve \mid I)}{P(D = paraguas \mid H = llueve, I) P(H = llueve \mid I) + P(D = paraguas \mid H = no\ llueve, I) P(H = no\ llueve \mid I)} = \frac{0.70 \times 0.20}{0.70 \times 0.20 + 0.10 \times 0.80} = 0.64$$

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

Resultado final:

$$P(H = no\ llueve \mid D = paraguas, I) = 0.36$$

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

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%) cabs.

¿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 = azul | I) = 0.15 **P(H = verde | I) = 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”

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

<u>Verosimilitud</u>

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

$$P(D = azul \mid H = verde, I) = 0.20$$

<u>Probabilidades a posteriori:</u>

$$P(H = azul \mid D = azul) = \frac{P(D = azul \mid H = azul, I) P(H = azul \mid I)}{P(D = azul \mid I)} = \frac{0.80 \times 0.15}{P(D = azul \mid I)} = \frac{0.12}{P(D = azul \mid I)}$$

$$P(H = verde \mid D = azul) = \frac{P(D = azul \mid H = verde, I) P(H = verde \mid I)}{P(D = azul \mid I)} = \frac{0.20 \times 0.85}{P(D = azul \mid I)} = \frac{0.17}{P(D = azul \mid I)}$$

Normalización:

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

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

Resultado final:

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

$$\mathbf{P(H} = verde \mid \mathbf{D} = azul, \mathbf{I}) = \mathbf{0.59} \quad \text{(solución MAP)}$$

Clasificando monedas como normales o trucadas

Consideremos una moneda que es lanzada (el resultado se guarda en la variable res).
Queremos decidir si la moneda es *normal* o está *trucada*.

¿Cuál es la probabilidad de errar si se sabe el resultado del lanzamiento y …

… decidimos moneda = Normal? $P(error) = P(moneda = Trucada \mid res)$

… decidimos moneda = Trucada? $P(error) = P(moneda = Normal \mid res)$

La decisión que lleva al **mínimo error** es predecir la clase que **maximiza la probabilidad a posteriori:**

Si $P\ (moneda = Normal \mid res) > P\ (moneda = Trucada \mid res)$ entonces clasificar como Normal

Si $P\ (moneda = Trucada \mid res) > P\ (moneda = Normal \mid res)$ entonces clasificar como Trucada

Ya que este problema de clasificación tiene dos clases:

$$P\ (moneda = Normal \mid res) + P\ (moneda = Trucada) \ = 1$$

Si $P\ (moneda = Normal \mid res)\ > 1/2$ entonces clasificar moneda como Normal

Si $P\ (moneda = Trucada \mid res) > 1/2$ entonces clasificar moneda como Trucada

Clasificando monedas como normales o trucadas

Supongamos lo siguiente:

Moneda normal: P (Cara) = 1/2 , P (Cruz) = 1/2

Moneda trucada: P (Cara) = 3/4 , P (Cruz) = 1/4

Tenemos una moneda y no sabemos si es normal o está trucada.

Realizamos una tirada y tratamos de adivinar si está trucada o no.

Hay cuatro posibles modelos de clasificación

Modelo A: Si sale Cara, decidir moneda = Trucada.
Si sale Cruz, decidir moneda = Normal.

Modelo B: Si sale Cara, decidir moneda = Normal.
Si sale Cruz, decidir moneda = Trucada.

Modelo C: Independientemente de lo que salga, decidir moneda = Normal.

Modelo D: Independientemente de lo que salga, decidir moneda = Trucada.

Sale Cara

Asumiendo que se lanza la moneda y sale Cara, ¿qué probabilidad hay de que sea una moneda normal?

$$P(moneda = Normal \mid res = Cara) = \frac{P(res = Cara \mid moneda = Normal)P(moneda = Normal)}{P(res = Cara)}$$

$$= \frac{1/2 \times P(moneda = Normal)}{P(res = Cara)}$$

$$P(moneda = Trucada \mid res = Cara) = \frac{P(res = Cara \mid moneda = Trucada)P(moneda = Trucada)}{P(res = Cara)}$$

$$= \frac{3/4 \times P(moneda = Trucada)}{P(res = Cara)}$$

Asumiendo prioris iguales $P(moneda = Trucada) = P(moneda = Normal) = 1/2$

$$P(moneda = Normal \mid res = Cara) = \frac{1/4}{P(res = Cara)} = 2/5 = 0.4$$

$$P(moneda = Trucada \mid res = Cara) = \frac{3/8}{P(res = Cara)} = 3/5 = 0.6$$

$$P(res = Cara) = 5/8 = 0.625$$

La clase que maximiza la probabilidad a posteriori cuando sale Cara es "Trucada"

Sale Cruz

Asumiendo que se lanza la moneda y sale Cruz, ¿qué probabilidad hay de que sea una moneda normal?

$$P(moneda = Normal \mid res = Cruz) = \frac{P(res = Cruz \mid moneda = Normal)P(moneda = Normal)}{P(res = Cruz)}$$

$$= \frac{1/2 \times P(moneda = Normal)}{P(res = Cruz)}$$

$$P(moneda = Trucada \mid res = Cruz) = \frac{P(res = Cruz \mid moneda = Trucada)P(moneda = Trucada)}{P(res = Cruz)}$$

$$= \frac{1/4 \times P(moneda = Trucada)}{P(res = Cruz)}$$

Asumiendo prioris iguales $P(moneda = Trucada) = P(moneda = Normal) = 1/2$

$$P(moneda = Normal \mid res = Cruz) = \frac{1/4}{P(res = Cruz)} = 2/3 = 0.67$$

$$P(moneda = Trucada \mid res = Cruz) = \frac{1/8}{P(res = Cruz)} = 1/3 = 0.33$$

$$P(res = Cruz) = 3/8 = 0.375$$

La clase que maximiza la probabilidad a posteriori cuando sale Cruz es “Normal”

Error del modelo

Prioris iguales: $P(moneda = Trucada) = P(moneda = Normal) = 1/2$

$$P(moneda = Normal \mid res = Cara) = 0.4 \qquad P(moneda = Trucada \mid res = Cara) = 0.6$$
$$P(moneda = Normal \mid res = Cruz) = 0.67 \qquad P(moneda = Trucada \mid res = Cruz) = 0.33$$
$$P(res = Cara) = 0.625 \qquad P(res = Cruz) = 0.375$$

¿Qué probabilidad de error tienen los diferentes modelos?

Modelo A: Si sale Cara, decidir moneda = Trucada. Si sale Cruz, decidir moneda = Normal.

$$\begin{aligned} P(error) &= P(res = Cruz, moneda = Trucada) + P(res = Cara, moneda = Normal) = \\ &= P(moneda = Trucada \mid res = Cruz)P(res = Cruz) + P(moneda = Normal \mid res = Cara)P(res = Cara) = \\ &= 0.33 \times 0.375 + 0.4 \times 0.625 = 0.37 \end{aligned}$$

Modelo B: Si sale Cara, decidir moneda = Normal. Si sale Cruz, decidir moneda = Trucada.

$$\begin{aligned} P(error) &= P(res = Cruz, moneda = Normal) + P(res = Cara, moneda = Trucada) = \\ &= P(moneda = Normal \mid res = Cruz)P(res = Cruz) + P(moneda = Trucada \mid res = Cara)P(res = Cara) = \\ &= 0.67 \times 0.375 + 0.6 \times 0.625 = 0.63 \end{aligned}$$

Modelo C: Decidir siempre moneda = Normal $\quad P(error) = P(moneda = Trucada) = 0.5$

Modelo D: Decidir siempre moneda = Trucada $\quad P(error) = P(moneda = Normal) = 0.5$

El mejor (minimiza el error): Modelo A.

Asigna a los ejemplos la clase que maximiza la probabilidad a posteriori.

Ejercicios

1. Hay 11 máquinas que fabrican circuitos electrónicos. 10 fabrican en promedio 1 circuito defectuoso de cada 6. La otra máquina es vieja y fabrica 1 defectuoso por cada 3.
   Los circuitos fabricados por las máquinas se almacenan en 11 cajas, una por máquina.
   Habiendo seleccionado una caja aleatoriamente, el primer circuito que examinamos es defectuoso. ¿Cuál es la probabilidad de que la caja corresponda a la máquina vieja?
   Asumamos ahora que se encuentran 2 circuitos defectuosos en extracciones consecutivas de una caja con un gran número de circuitos.
   ¿Cuál es la probabilidad de que corresponda a la máquina vieja?

2. En una caja tenemos monedas trucadas (75% probabilidad de salir cara) y monedas normales (50% probabilidad de salir cara). Tomamos una al azar, la lanzamos y sale cara. ¿Cuál debe ser la proporción de monedas en la caja para que este resultado no nos ayude a decidir si la moneda está trucada o no?
   Nota que en este caso la probabilidad a posteriori parece darnos menos información que el priori.

3. Un estudio estadístico muestra que el 70% de los ciclistas en una ciudad son hombres.
   De ellos, el 60% llevan casco.
   El porcentaje de mujeres que llevan casco cuando van en bici es un 40%.
   ¿Cuál es la probabilidad de que un ciclista seleccionado al azar lleve casco?
   Si un/a ciclista seleccionad@ al azar lleva casco,
   ¿cuál es la probabilidad de que sea hombre?

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'

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

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

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

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

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

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

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

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

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’

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

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

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

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

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

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

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

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

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)**

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

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

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

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

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)**

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

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

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

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

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)
miope	*normal*	*duras*	*duras* [50%]
miope	*reducido*	*no*	*no* [100%]
hipermétrope	*normal*	*blandas*	*blandas* [50%]
hipermétrope	*reducido*	*no*	*no* [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).**

Clasificador Naïve Bayes

- **Independencia condicional**

A es condicionalmente independiente de B dado C si

$P(A \mid B, C) = P(A \mid C)$ (ecuación 1)

Que es equivalente a:

$P(A, B \mid C) = P(A \mid C) \cdot P(B \mid C)$ (ecuación 2)

Dem.

Usamos la regla del producto: $P(A, B \mid C) = P(A \mid B, C) \cdot P(B \mid C)$ ("ecuación 3")

Igualando ecs. 2 y 3 obtenemos $P(A \mid B, C) \cdot P(B \mid C) = P(A \mid C) \cdot P(B \mid C)$

Lo que significa que ec. 2 implica la ec. 1, y que ec. 1 implica la ec. 2 (si multiplicamos a ambos lados la ec. 1 por $P(B|C)$ y usamos la ec. 3, obtenemos la ec. 2)

- **Dado un problema de clasificación**

Atributos: $\mathbf{x} = \{x^{(1)}, x^{(2)}, \ldots, x^{(D)}\}$

Clase : $c \in \{1,2,\ldots,C\}$

Naïve Bayes: Asume que los atributos son condicionalmente independientes dada la clase

$$P(c \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid c)P(c)}{P(\mathbf{x})} = \frac{P(x^{(1)}, x^{(2)}, \ldots, x^{(D)} \mid c)P(c)}{P(\mathbf{x})} \approx \frac{P(x^{(1)} \mid c)P(x^{(2)} \mid c)\ldots P(x^{(D)} \mid c)P(c)}{\sum_{c'=1}^{C} P(x^{(1)} \mid c')P(x^{(2)} \mid c')\ldots P(x^{(D)} \mid c')P(c')}$$

Clasificador Naïve Bayes

- **Ejemplo 1 del problema de las lentillas: prescr. =‘miope’, lagr. = ‘normal’**
- **Prioris de la clase** $P(c='n')=\frac{15}{24}\quad P(c='b')=\frac{5}{24}\quad P(c='d')=\frac{4}{24}$
- **Marginales condicionados:**

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

$$\boxed{\begin{aligned}P(p='m'|c='n')&=\frac{7}{15}\\ P(p='m'|c='b')&=\frac{2}{5}\\ P(p='m'|c='d')&=\frac{3}{4}\end{aligned}}$$

	no	blandas	duras
t=‘normal’	3	5	4
t=‘reducido’	12	0	0

$$\boxed{\begin{aligned}P(l='n'|c='n')&=\frac{3}{15}\\ P(l='n'|c='b')&=\frac{5}{5}\\ P(l='n'|c='d')&=\frac{4}{4}\end{aligned}}$$

- **Suposición de independencia condicional:**

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

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

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

- **Naïve Bayes:**

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

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

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

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

Estimador de Laplace

- **Estimación de probabilidades con frecuencias:**

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

- **Estimador de Laplace:** Añadir ejemplos ficticios

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

  – Evita estimaciones nulas para las probabilidades
  – Estimaciones más robustos
  – Asintóticamente pequeño

	no	blandas	duras
t='normal'	3+1	5+1	4+1
t='reducido'	12+1	0+1	0+1

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

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

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

Redes Bayesianas (“Modelos gráficos”)

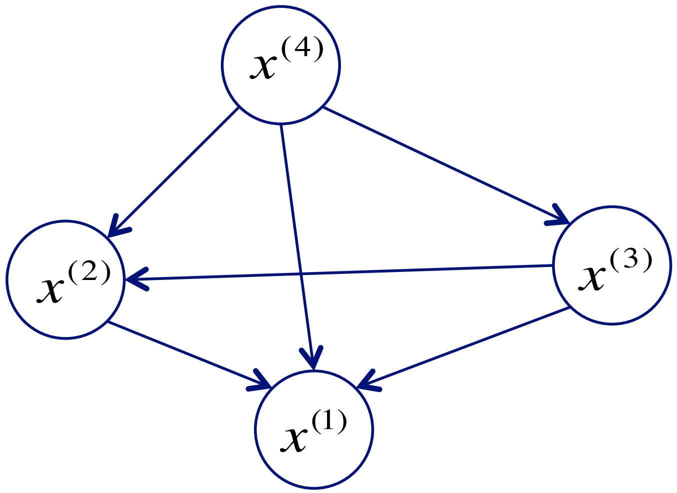
Regla de la cadena

$$P(x^{(1)},x^{(2)},x^{(3)},x^{(4)}) = P(x^{(1)} \mid x^{(2)},x^{(3)},x^{(4)})P(x^{(2)},x^{(3)},x^{(4)}) =$$
$$= P(x^{(1)} \mid x^{(2)},x^{(3)},x^{(4)})P(x^{(2)} \mid x^{(3)},x^{(4)})P(x^{(3)},x^{(4)}) =$$
$$= P(x^{(1)} \mid x^{(2)},x^{(3)},x^{(4)})P(x^{(2)} \mid x^{(3)},x^{(4)})P(x^{(3)} \mid x^{(4)})P(x^{(4)})$$

Representación gráfica de la regla de la cadena:
Grafo acíclico dirigido que representa las dependencias entre las variables

- **Nodos:** Variables
- **Aristas dirigidas:** Dependencias

Una flecha hacia un nodo dado corresponde a variables dependientes de la variable que origina la flecha



Nota: La representación gráfica **no es única!**

Interpretación del grafo:

Nodo i: Variable $x^{(i)}$, $i = 1,2,\ldots,n$
$\Pi(x^{(i)})$: Padres del nodo $x^{(i)}$

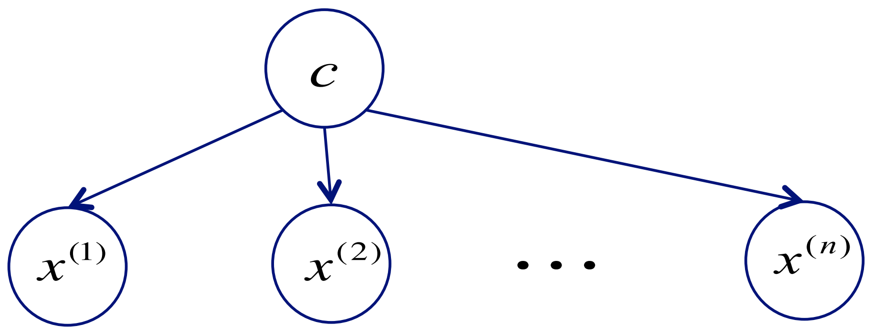
$$P(x^{(1)},x^{(2)},\ldots,x^{(n)}) = \prod_{i=1}^{n} P\left(x^{(i)} \mid \Pi(x^{(i)})\right)$$

Nota: Si $\Pi(x^{(i)}) = \varnothing$ entonces $P(x^{(i)} / \Pi(x^{(i)})) = P(x^{(i)})$

Naïve Bayes

$$P(x^{(1)},x^{(2)},\ldots,x^{(n)},c) = \prod_{i=1}^{n} P\left(x^{(i)} \mid c\right) P(c)$$

Modelo factorizado



Problema de las lentes de contacto

- **Ejemplo: se observa edad='joven', astigmatismo='sí', lagrimeo = 'normal'**
- **Prioris de las clases** $P(c='n')=\frac{15}{24}\quad P(c='b')=\frac{5}{24}\quad P(c='d')=\frac{4}{24}$
- **Marginales condicionados:**

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

$$P(e='m'|c='n')=\frac{4}{15}\qquad P(e='m'|c='b')=\frac{2}{5}\qquad P(e='m'|c='d')=\frac{2}{4}$$

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

$$P(as='sí'|c='n')=\frac{4}{15}\qquad P(as='sí'|c='b')=\frac{2}{5}\qquad P(as='sí'|c='d')=\frac{2}{4}$$

	no	blandas	duras
t='normal'	3	5	4
Total	15	5	4

$$P(l='n'|c='n')=\frac{3}{15}\qquad P(l='n'|c='b')=\frac{5}{5}\qquad P(l='n'|c='d')=\frac{4}{4}$$

- **Naive Bayes.**

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

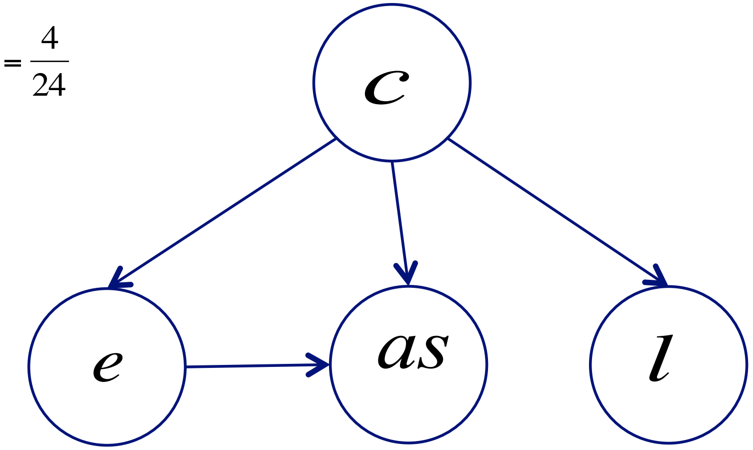
$$P(c='b'|e='j',as='sí',l='n')=\frac{P(e='j',as='sí',l='n'|c='b')P(c='b')}{P(e='j',as='sí',l='n')}\approx\frac{\frac{2}{5}\times\frac{0}{5}\times\frac{5}{5}\times\frac{5}{24}}{Norm}=0.00\quad[0\%]$$

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

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

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

$$P(e='j'|c='n')=\frac{4}{15}$$
$$P(e='j'|c='b')=\frac{2}{5}$$
$$P(e='j'|c='d')=\frac{2}{4}$$

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

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

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

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