Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Sistemas Inteligentes

Tema 5.2: Redes Bayesianas

Curso 2024-25

Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Índice

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Teorema de Bayes

- Sabemos que:
  - P(A|B) P(B) = P(A,B)
  - P(B|A) P(A) = P(B,A) = P(A,B)
- Regla de Bayes:

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

- Constante de normalización P(B)

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

Marginalización
Prob. total

- Regla de la cadena:

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

$$P(A,B,C) = P(A)P(B|A)P(C|B,A)$$

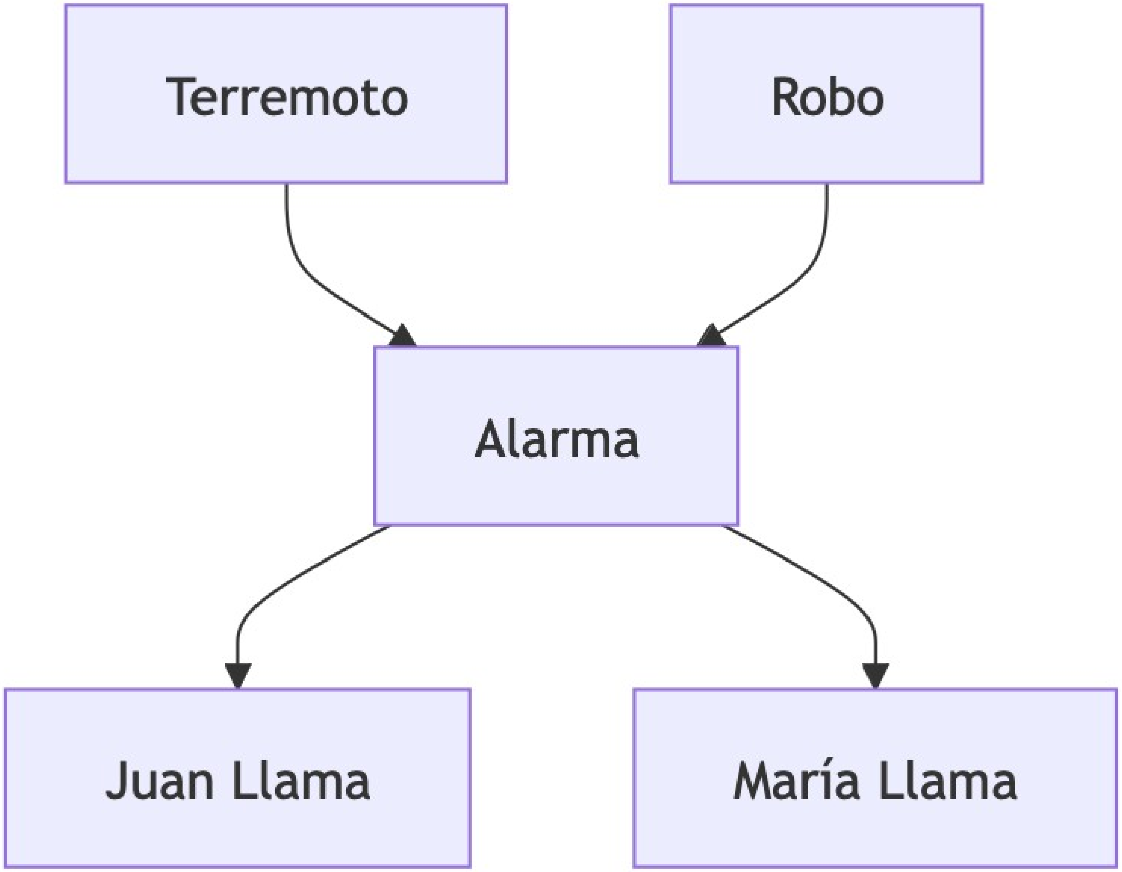
Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas

- Una red bayesiana es:
  - Un grafo acíclico dirigido para representar dependencias entre variables y mostrar una descripción escueta de cualquier distribución de probabilidad conjunta completa
- Esta formada por
  - Un conjunto de variables aleatorias que forman los nodos de la red. Cada nodo X tendrá adjunta una distribución P(X|Padres(X))
  - Un conjunto de enlaces que determinan la influencia (dependencia) entre nodos. Si X se conecta con Y se dice que X influencia a Y
- Su finalidad principal es calcular la distribución conjunta de las variables nodo

Semántica de una red

- Dada la siguiente red bayesiana...
  - ¿Qué distribución representa?

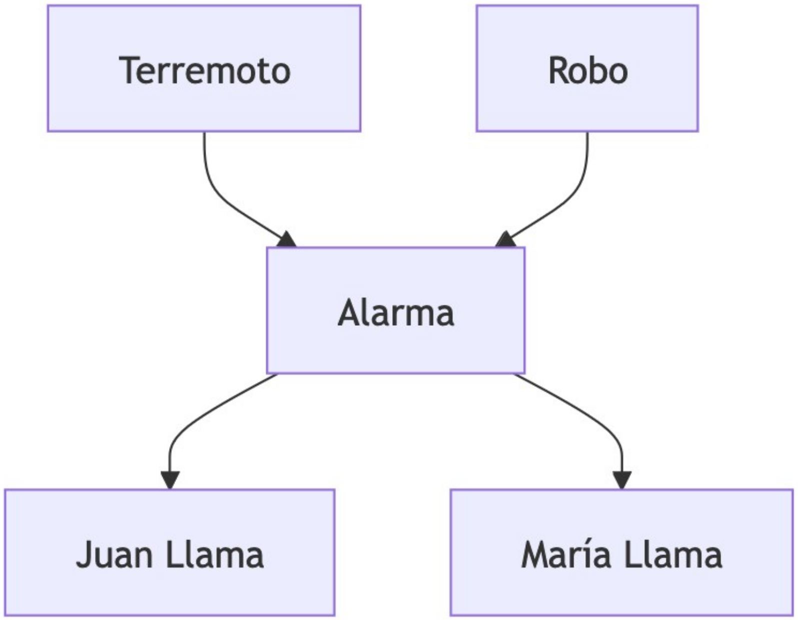


Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Semántica de una red

- Dada la siguiente red bayesiana...
  - P(T,R,A,J,M) =

    P(T) * P(R) * P(A|T,R) * P(J|A) * P(M|A)
  - ¿ |P(T,R,A,J,M)| sin independencia condicional?
    - 2^5 = 32
  - ¿Y con independencia condicional?
    - 2+2+2^3+2^2 +2^2 = 20

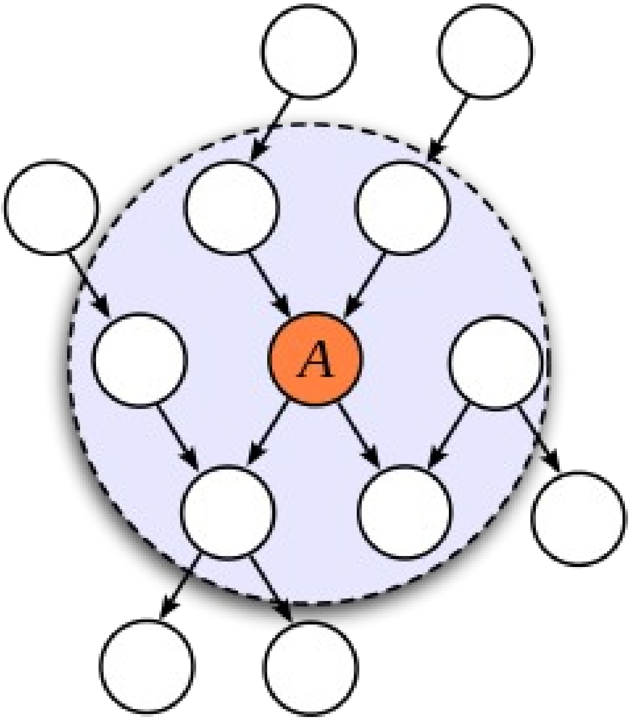


Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Cobertura de Markov

- Cobertura de Markov
  - Un nodo A es condicionalmente independiente de todos los nodos de la red dados:
    - Sus padres
    - Sus hijos
    - Los padres de sus hijos
  - Nos ayuda a:
    - Identificación de independencias condicionales
    - Simplificación del modelo
    - Mejorar la inferencia
    - Interpretación de relaciones
    - Reducción de complejidad



Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Información Mutua

Recordemos sobre la ganancia de Información:

- Mide la reducción en la entropía de la variable objetivo (clase) al conocer el valor de un atributo.
- GI(Clase, Atributo) = E(Clase) - E(Clase|Atributo)
- **Es un caso especial de Información Mutua,** utilizado comúnmente en árboles de decisión y selección de características.

Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Información Mutua

Información Mutua (IM):

- Estrechamente relacionada con la Ganancia de Información
- Mide la reducción en la incertidumbre de una variable aleatoria debido al conocimiento de otra.
- Es simétrica: I(X;Y) = I(Y;X)
- Se define como: I(X;Y) = E(X) - E(X|Y) = E(Y) – E(Y|X)
- Lo usaremos para determinar que nodo es más influyente respecto a otro

$$\boxed{\begin{aligned} I(X;Y) &= E(X) - E(X|Y) \\ E(X|Y) &= -\sum_{x,y} p(x,y)\log p(x|y) \\ I(X;Y) &= \sum_{x,y} P(x,y)\cdot \log_2\left(\frac{P(x,y)}{P(x)\cdot P(y)}\right) \end{aligned}}$$

Notar que: la entropía cond. es consistente con el tema anterior:

$$\begin{aligned} E(Y|X) &= \sum_i P(X=x_i)\cdot E(Y|X=x_i) \\ &= \sum_i P(X=x_i)\cdot\left[-\sum_j P(Y=y_j|X=x_i)\log P(Y=y_j|X=x_i)\right] \\ &= -\sum_i\sum_j P(X=x_i)\cdot P(Y=y_j|X=x_i)\log P(Y=y_j|X=x_i) \\ &= -\sum_i\sum_j P(X=x_i,Y=y_j)\log P(Y=y_j|X=x_i) \\ &= -\sum_{x,y} p(x,y)\log p(y|x) \\ E(X|Y) &= -\sum_{x,y} p(x,y)\log p(x|y) \end{aligned}$$

Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia

- ¿Para que queremos la distribución conjunta?
  - A partir de la distribución conjunta podemos contestar cualquier pregunta relativa a la red…
- Varios tipos de inferencia en redes Bayesianas
  - Exacta (caso general)
  - Casos especiales (Kim&Pearl…)
  - Aproximada

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia exacta

- Inferencia exacta general (funciona para todas la RR.BB.)
- Regla de inferencia general
  - (Donde B son las variables buscadas, C las conocidas y D las desconocidas)

$$P(B|C) = \alpha \cdot \sum_{D} P(B, D, C)$$

- Problema: mucha complejidad

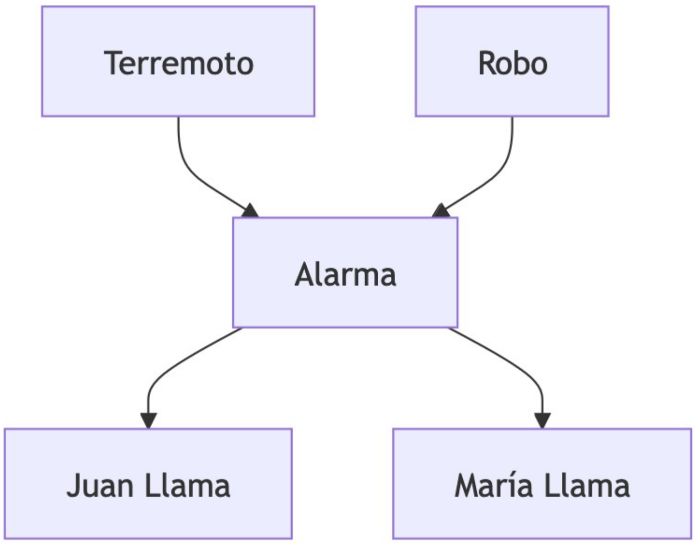
Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Inferencia exacta: ejemplo 1

- ¿Prob. que suene la alarma si llama María?

$$P(B|C) = \alpha \cdot \sum_{D} P(B, D, C)$$

- P(R,T,A,J,M)= P(R)*P(T)*P(A|R,T)* P(J|A)*P(M|A)



Universitat d'Alacant
Universidad de Alicante

Inferencia exacta: ejemplo 1

- De esta manera tenemos que:

$$P(A|M) = \alpha \cdot \sum_R \sum_T \sum_J P(R,T,A,J,M) =$$

$$= \alpha \cdot \sum_R \sum_T \sum_J P(R) \cdot P(T) \cdot P(A|R,T) \cdot P(J|A) \cdot P(M|A) =$$

$$= \alpha \cdot P(M|A) \cdot \sum_R \left(P(R) \sum_T \left(P(T) \cdot P(A|R,T) \cdot \underbrace{\sum_J P(J|A)}_{1} \right) \right)$$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Inferencia exacta: ejemplo 2

- **¿P(R|J+,M+)?**

P(J|A) =

A	J
0	0,05
1	0,9

P(M|A) =

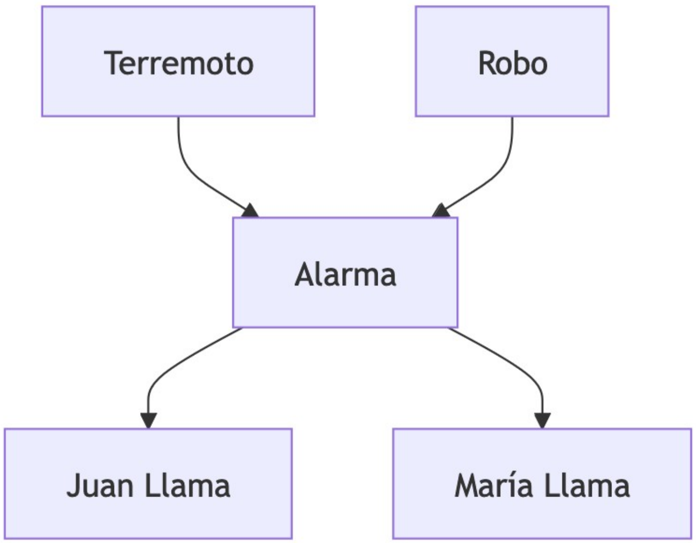
A	M
0	0,01
1	0,7

P(A|T,R)=

T	R	A
0	0	0,001
0	1	0,94
1	0	0,29
1	1	0,95

P(T) = 0,001

P(R) = 0,002



Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia exacta: ejemplo 2

- De esta manera tenemos que:

$$P(R|J,M) = \alpha \sum_T \sum_A P(R,T,A,J,M) =$$

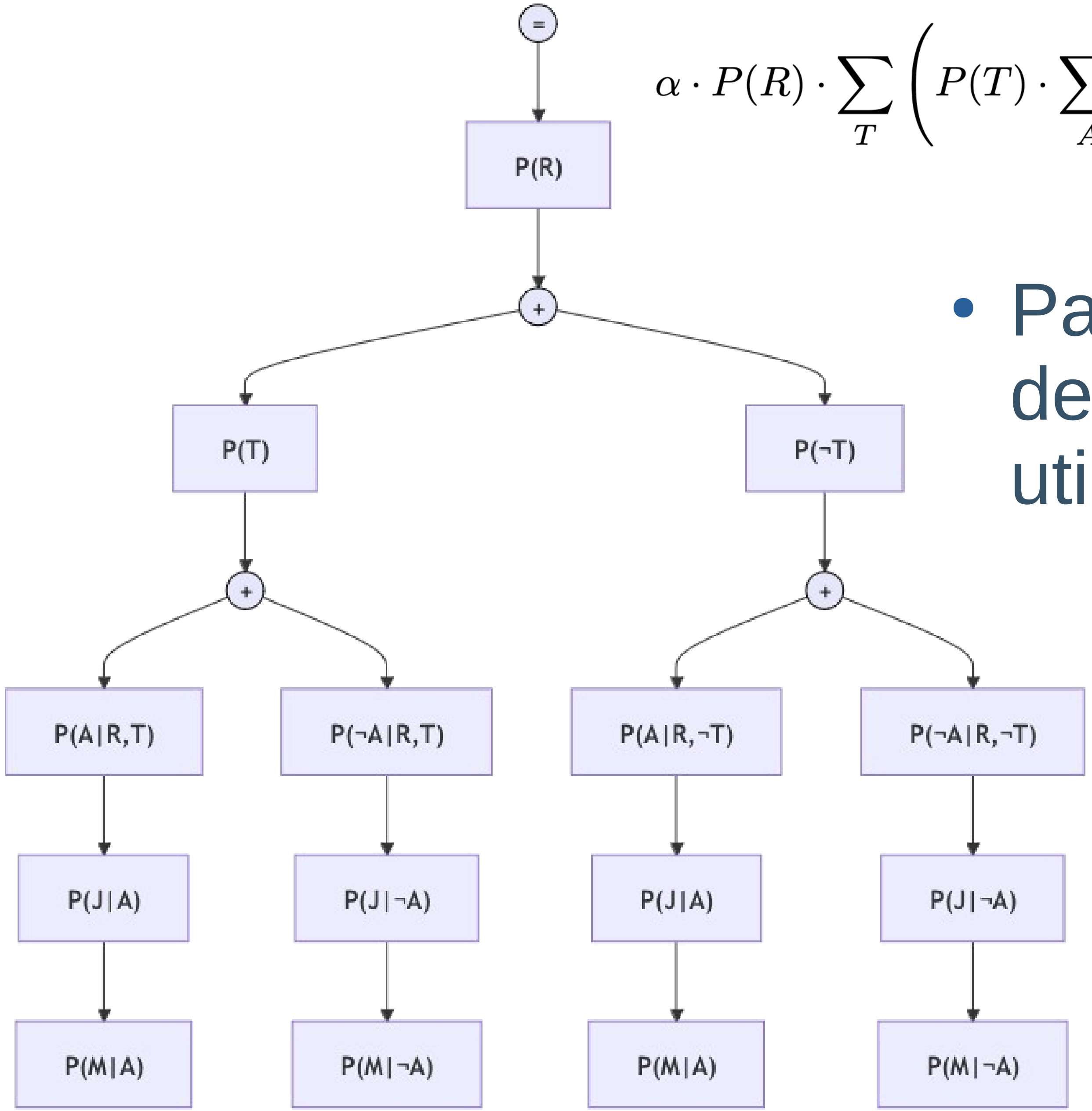
$$\alpha \sum_T \sum_A P(R) \cdot P(T) \cdot P(A|R,T) \cdot P(J|A) \cdot P(M|A) =$$

$$\alpha \cdot P(R) \cdot \sum_T \left(P(T) \cdot \sum_A (P(A|R,T) \cdot P(J|A) \cdot P(M|A)) \right)$$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia exacta: ejemplo 2

$$\alpha \cdot P(R) \cdot \sum_{T} \left(P(T) \cdot \sum_{A} \left(P(A|R,T) \cdot P(J|A) \cdot P(M|A)\right)\right)$$

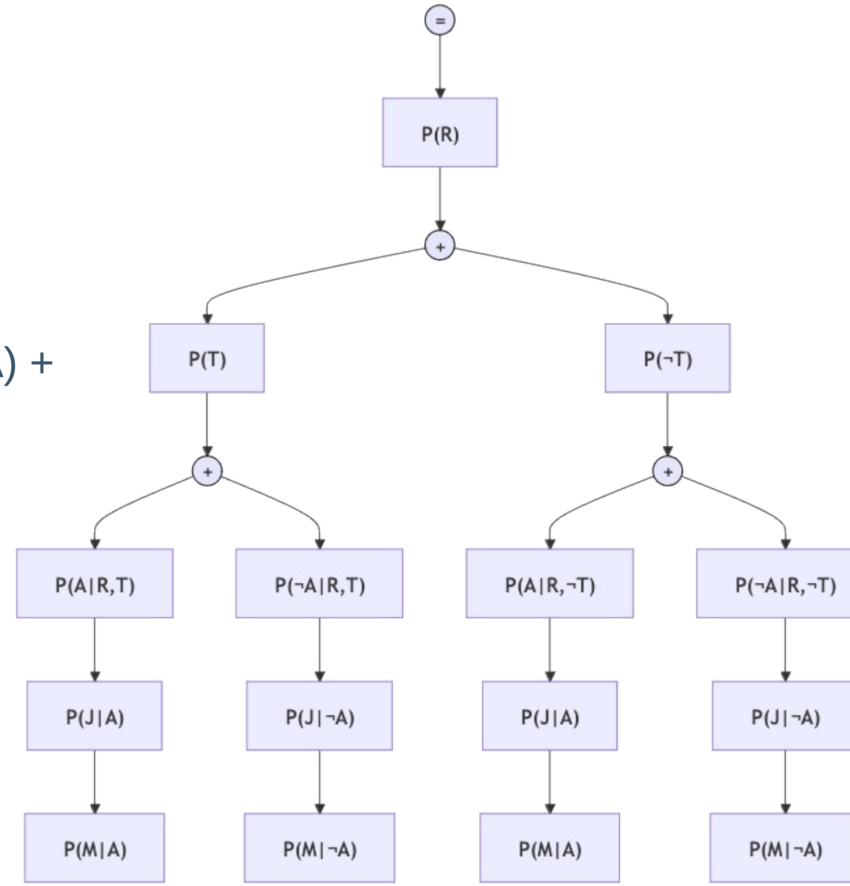


- Para calcular descomponemos utilizando un árbol...

Inferencia exacta: ejemplo 2

- $P(R|J,M) = \boldsymbol{\alpha} \cdot P(R) \cdot \Sigma_T (P(T) \cdot \Sigma_A (P(A|R,T) \cdot P(J|A) \cdot P(M|A)))$

- **P(R|J,M)** $= \alpha * P(R) * [P(T) * P(A|R,T) * P(J|A) * P(M|A) +$
  $P(T) * P(\neg A|R,T) * P(J|\neg A) * P(M|\neg A) +$
  $P(\neg T) * P(A|R,\neg T) * P(J|A) * P(M|A) +$
  $P(\neg T) * P(\neg A|R,\neg T) * P(J|\neg A) * P(M|\neg A)]$
- **P(¬R|J,M)** $= \alpha * P(\neg R) * [P(T) * P(A|\neg R,T) * P(J|A) * P(M|A) +$
  $P(T) * P(\neg A|\neg R,T) * P(J|\neg A) * P(M|\neg A) +$
  $P(\neg T) * P(A|\neg R,\neg T) * P(J|A) * P(M|A) +$
  $P(\neg T) * P(\neg A|\neg R,\neg T) * P(J|\neg A) * P(M|\neg A)]$
- **P(R|J,M) + P(¬R|J,M)** $= 1 \rightarrow \alpha = 1/(P(R|J,M)+P(\neg R|J,M))$



Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia exacta: ejemplo 2

- Cálculo de P(R|J,M):
  - **P(R|J,M)** = α* P(R) *
    [P(T) * P(A|R,T) * P(J|A) * P(M|A) +
    P(T) * P(¬A|R,T) * P(J|¬A) * P(M|¬A) +
    P(¬T) * P(A|R,¬T) * P(J|A) * P(M|A) +
    P(¬T) * P(¬A|R,¬T) * P(J|¬A) * P(M|¬A)]
  - P(A|R,T) * P(J|A) * P(M|A) = 0,95 * 0,9 * 0,7 = 0,5985
  - P(¬A|R,T) * P(J|¬A) * P(M|¬A) = 0,05 * 0,05 * 0,01 =0,000025
  - P(A|R,¬T) * P(J|A) * P(M|A) = 0,94 * 0,9 * 0,7 = 0,5922
  - P(¬A|R,¬T) * P(J|¬A) * P(M|¬A) = (1-0,94)* 0,05 * 0,01 = 0,00003
  - P(R|J,M) = α * P(R) * [P(T) * (0,5985 + 0,000025) + P(¬T) * (0,5922 + 0,00003)]
    = 0,002 * [0,001 * (0,5985 + 0,000025) + 0,999 * (0,5922 + 0,00003)]
    = 0,002 * [0,000598525 + 0,59223]
    = 0,002 * 0,592828525
    = 0,00118565705

P(J|A) =

A	J
0	0,05
1	0,9

P(A|T,R)=

T	R	A
0	0	0,001
0	1	0,94
1	0	0,29
1	1	0,95

P(T) = 0,001

P(R) = 0,002

P(M|A) =

A	M
0	0,01
1	0,7

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Inferencia exacta: ejemplo 2

- Cálculo de P(R|J,M):
  - P(R|J,M) = α * P(R) * [P(T) * (0,5985 + 0,000025) + P(¬T) * (0,5922 + 0,00003)]
    = 0,002 * [0,001 * (0,5985 + 0,000025) + 0,999 * (0,5922 + 0,00003)]
    = 0,002 * [0,000598525 + 0,59223]
    = 0,002 * 0,592828525
    = 0,00118565705
- Cálculo de P(¬R|J,M):
  - P(¬R|J,M) = α * P(¬R) * [P(T) * (0,1827 + 0,000355) + P(¬T) * (0,00063 + 0,0004995)]

    = 0,998 * [0,001 * (0,1827 + 0,000355) + 0,999 * (0,00063 + 0,0004995)]

    = 0,998 * [0,000183055 + 0,00112935]

    = 0,998 * 0,001312405

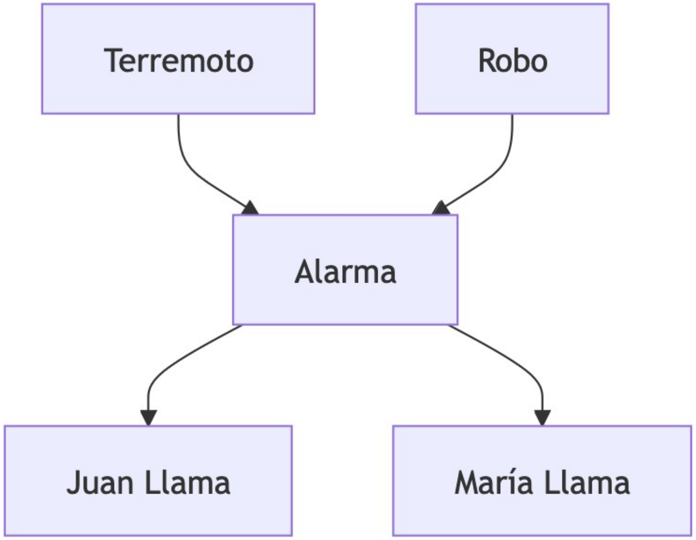
    = 0,00130997819
- Probabilidad total
  - P(R|J,M) + P(¬R|J,M) = 0,00118565705 + 0,00130997819
  - α = 1 / 0,00249563524 = 400,70003
- Probabilidades finales:
  - P(R|J,M) = α * P(R|J,M) = 400,70003 * 0,00118565705 = 0,47509
  - P(¬R|J,M) = α * P(¬R|J,M) = 400,70003 * 0,00130997819 = 0,52491

Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia exacta: ejercicio

- Ejercicio: ¿P(J|R)?



$$P(J|R) = \sum_T \sum_A \sum_M P(R) \cdot P(T) \cdot P(A|R,T) \cdot P(J|A) \cdot P(M|A) =$$

$$= P(R) \cdot \sum_T \left(P(T) \cdot \sum_A \left(P(A|R,T) \cdot P(J|A) \cdot \sum_M P(M|A) \right) \right)$$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Influencia entre nodos

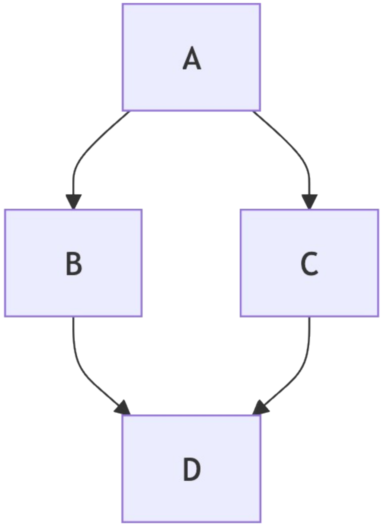
- Usamos la información mutua para comparar entre nodos
- Ejemplo. Dada la siguiente red

A	P(A)
True	0.5
False	0.5

A	P(C = True | A)	P(C = False | A)
True	0.7	0.3
False	0.4	0.6

A	P(B = True | A)	P(B = False | A)
True	0.8	0.2
False	0.3	0.7

B	C	P(D = True | B,C)	P(D = False | B,C)
True	True	0.9	0.1
True	False	0.7	0.3
False	True	0.6	0.4
False	False	0.1	0.9



- ¿Qué nodo influencia más a D?

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Influencia entre nodos

- Calculemos la información mutua I(X;D) para X = A, B, y C:
  - $$I(X;Y) = \sum_{x,y} P(x,y) \cdot \log_2 \left(\frac{P(x,y)}{P(x) \cdot P(y)} \right)$$
  - Primero, necesitamos calcular las probabilidades marginales:
    - P(B+) = P(A)*P(B|A) + P(A-)*P(B|A-)
      = 0.5 * 0.8 + 0.5 * 0.3 = 0.55;
    - P(C+) = P(A)*P(C|A) + P(A-)*P(C|A-)
      = 0.5 * 0.7 + 0.5 * 0.4 = 0.55;
    - P(D+) = P(B)*P(C)*P(D|B,C) +
      P(B)*P(C-)*P(D|B,C-)+
      P(B-)*P(C)*P(D|B-,C)+
      P(B-)*P(C-)*P(D|B-,C-)=

      0.55 * 0.55 * 0.9 +
      0.55 *0.45 *0.7 +
      0.45 *0.55 *0.6 +
      0.45 *0.45 *0.1 ≈ 0.6

A	P(A)
True	0.5
False	0.5

A	P(C = True | A)	P(C = False | A)
True	0.7	0.3
False	0.4	0.6

A	P(B = True | A)	P(B = False | A)
True	0.8	0.2
False	0.3	0.7

B	C	P(D = True | B,C)
True	True	0.9
True	False	0.7
False	True	0.6
False	False	0.1

Influencia entre nodos

- Calculemos la información mutua I(X;D) para X = A, B, y C:
  - $$I(X;Y) = \sum_{x,y} P(x,y) \cdot \log_2 \left(\frac{P(x,y)}{P(x) \cdot P(y)} \right)$$
  - Primero, necesitamos calcular las probabilidades marginales:
    - P(B+) = 0.5 * 0.8 + 0.5 * 0.3 = 0.55; P(C+) = 0.5 * 0.7 + 0.5 * 0.4 = 0.55; P(D+) = 0.55 * 0.55 * 0.9 + 0.55 *0.45 *0.7 + 0.45 *0.55 *0.6 + 0.45 *0.45 *0.1= 0.6
  - Ahora, calculemos I(X;**D**) para cada X:
    - Para A:
      - I(A;D) =
        P(A+,D+)*log2(P(A+,D+)/(P(A+)*P(D+))) +
        P(A+,D-)*log2(P(A+,D-)/(P(A+)*P(D-))) +
        P(A-,D+)*log2(P(A-,D+)/(P(A-)*P(D+))) +
        P(A-,D-)*log2(P(A-,D-)/(P(A-)*P(D-)))
      - ¿P(A,D)? =
        - P(A,D) = P(D,A) = P(D|A)P(A)

$$P(D,A) = \sum_B \sum_C P(A,B,C,D)$$

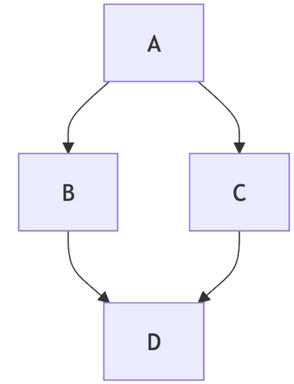
$$= \sum_B \sum_C P(D|B,C) * P(B|A) * P(C|A) * P(A)$$

A	P(A)
True	0.5
False	0.5

A	P(C = True | A)	P(C = False | A)
True	0.7	0.3
False	0.4	0.6

A	P(B = True | A)	P(B = False | A)
True	0.8	0.2
False	0.3	0.7

B	C	P(D = True | B,C)
True	True	0.9
True	False	0.7
False	True	0.6
False	False	0.1



Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Influencia entre nodos

- P(D,A)

$$P(A,D) = \sum_B \sum_C P(A,B,C,D)$$
$$= \sum_B \sum_C P(D|B,C) * P(B|A) * P(C|A) * P(A)$$

= [P(D+|B+,C+)P(B+|A)P(C+|A) +
P(D+|B+,C-)P(B+|A)P(C-|A) +
P(D+|B-,C+)P(B-|A)P(C+|A) +
P(D+|B-,C-)P(B-|A)P(C-|A)] * P(A)

A	P(A)
True	0.5
False	0.5

A	P(C = True | A)	P(C = False | A)
True	0.7	0.3
False	0.4	0.6

A	P(B = True | A)	P(B = False | A)
True	0.8	0.2
False	0.3	0.7

B	C	P(D = True | B,C)
True	True	0.9
True	False	0.7
False	True	0.6
False	False	0.1

Para A+:

P(D+,A+) = 0.5 * [
0.9 * 0.8 * 0.7 + // B+,C+
0.7 * 0.8 * 0.3 + // B+,C-
0.6 * 0.2 * 0.7 + // B-,C+
0.1 * 0.2 * 0.3] // B-,C-
= 0.5 * [0.504 + 0.168 + 0.084 + 0.006] = 0.381

P(A+,D+) = 0,381; P(A+) = P(A+,D+) **+** P(A+,D-)

P(A+,D-) = P(A+) - P(A+,D+) = 0.119

Para A-:

P(D+,A-) = 0.5 * [
0.9 * 0.3 * 0.4 + // B+,C+
0.7 * 0.3 * 0.6 + // B+,C-
0.6 * 0.7 * 0.4 + // B-,C+
0.1 * 0.7 * 0.6] // B-,C-
= 0.5 * [0.108 + 0.126 + 0.168 + 0.042] = 0.22

P(A-,D+) = 0.222; P(A+) = P(A+,D+) **+** P(A+,D-)

P(A-,D-) = P(A-) - P(A-,D-) = 0.278

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Influencia entre nodos

- Calculemos la información mutua I(X;D) para X = A, B, y C:
  - $$I(X;Y) = \sum_{x,y} P(x,y) \cdot \log_2 \left(\frac{P(x,y)}{P(x) \cdot P(y)} \right)$$
  - Primero, necesitamos calcular las probabilidades marginales:
    - P(B+) = 0.5 * 0.8 + 0.5 * 0.3 = 0.55; P(C+) = 0.5 * 0.7 + 0.5 * 0.4 = 0.55;

      P(D+) = 0.55 * 0.55 * 0.9 + 0.55 *0.45 *0.7 + 0.45 *0.55 *0.6 + 0.45 *0.45 *0.1 ≈ 0.6
  - Ahora, calculemos I(X;**D**) para cada X:
    - Para A:
      - I(A;D) =
        P(A+,D+)*log2(P(A+,D+)/(P(A+)*P(D+))) +
        P(A+,D-)*log2(P(A+,D-)/(P(A+)*P(D-))) +
        P(A-,D+)*log2(P(A-,D+)/(P(A-)*P(D+))) +
        P(A-,D-)*log2(P(A-,D-)/(P(A-)*P(D-)))
      - P(A+) = 0.5, P(A-) = 0.5
      - I(A;D) =

        0.381 * log2(0.381/(0.5*0.6)) + 0.119 * log2(0.119/(0.5*0.4)) +

        0.222 * log2(0.222/(0.5*0.6)) + 0.278 * log2(0.278/(0.5*0.4))

        ≈ 0.078 bits

A	P(A)
True	0.5
False	0.5

A	P(C = True | A)	P(C = False | A)
True	0.7	0.3
False	0.4	0.6

A	P(B = True | A)	P(B = False | A)
True	0.8	0.2
False	0.3	0.7

B	C	P(D = True | B,C)
True	True	0.9
True	False	0.7
False	True	0.6
False	False	0.1

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Influencia entre nodos

- Ahora, calculemos I(X;**D**) para cada X:
  - **Para B:**

    P(B+,D+) = 0.55*0.824 ≈ 0.453; P(B+,D-) = 0.55*0.176 ≈ 0.097

    P(B-,D+) = 0.45*0.333 ≈ 0.15; P(B-,D-) = 0.45*0.667 ≈ 0.3

    I(B;D) = 0.453*log2(0.453/(0.55*0.6)) +

    0.097*log2(0.097/(0.55*0.4)) +

    0.15*log2(0.15/(0.45*0.6)) +

    0.3*log2(0.3/(0.45*0.4))

    ≈ 0.186 bits
  - **Para C:**

    P(C+,D+) = 0.55*0.785 ≈ 0.432; P(C+,D-) = 0.55*0.215 ≈ 0.118

    P(C-,D+) = 0.45*0.38 ≈ 0.171; P(C-,D-) = 0.45*0.62 ≈ 0.279

    I(C;D) = 0.432*log2(0.432/(0.55*0.6)) +

    0.118*log2(0.118/(0.55*0.4)) +

    0.171*log2(0.171/(0.45*0.6)) +

    0.279*log2(0.279/(0.45*0.4))

    ≈ 0.125 bits
- La información mutua más alta es para B, lo que indica que B tiene la relación más fuerte con D en términos de información compartida.

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Ejercicio: diseña una red bayesi.

- Sistema de Predicción de Éxito Académico
  - Una universidad está desarrollando un sistema para predecir el éxito académico de sus estudiantes. Se han identificado los siguientes factores que podrían influir en el rendimiento de un estudiante:
    - Hábitos de estudio (buenos o malos)
    - Asistencia a clases (regular o irregular)
    - Participación en actividades extracurriculares (sí o no)
    - Nivel de estrés (alto o bajo)
    - Rendimiento académico (alto o bajo)
- Se cree que los hábitos de estudio y la asistencia a clases influyen directamente en el rendimiento académico. La participación en actividades extracurriculares puede afectar tanto al nivel de estrés como a los hábitos de estudio. El nivel de estrés, a su vez, puede influir en la asistencia a clases y en el rendimiento académico.

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Ejercicio: diseña una red bayesi.

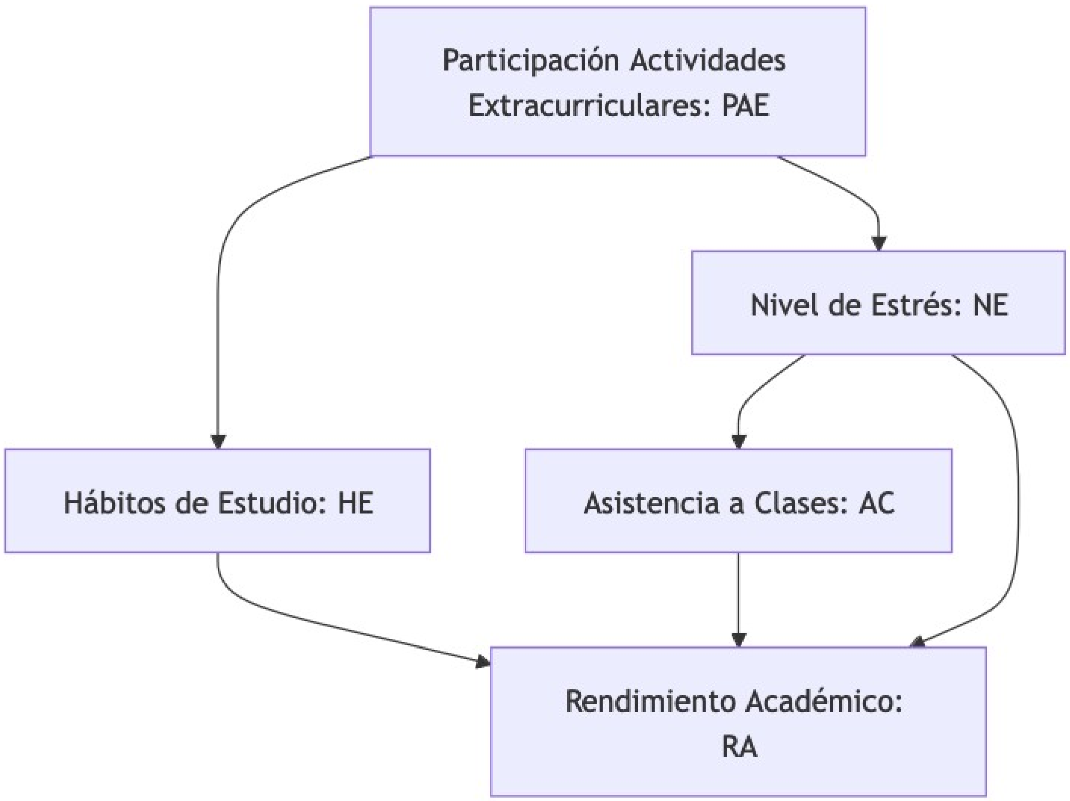
- Realiza los siguientes apartados:
  - Diseña la red
  - Calcula que distribución de prob. representa
  - Calcular la probabilidad de que un estudiante tenga un alto rendimiento académico dado que tiene buenos hábitos de estudio y baja asistencia a clases.
  - Determinar qué factor tiene mayor impacto en el rendimiento académico según el modelo creado

Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Ejercicio: diseña una red bayesi.

- Solución:

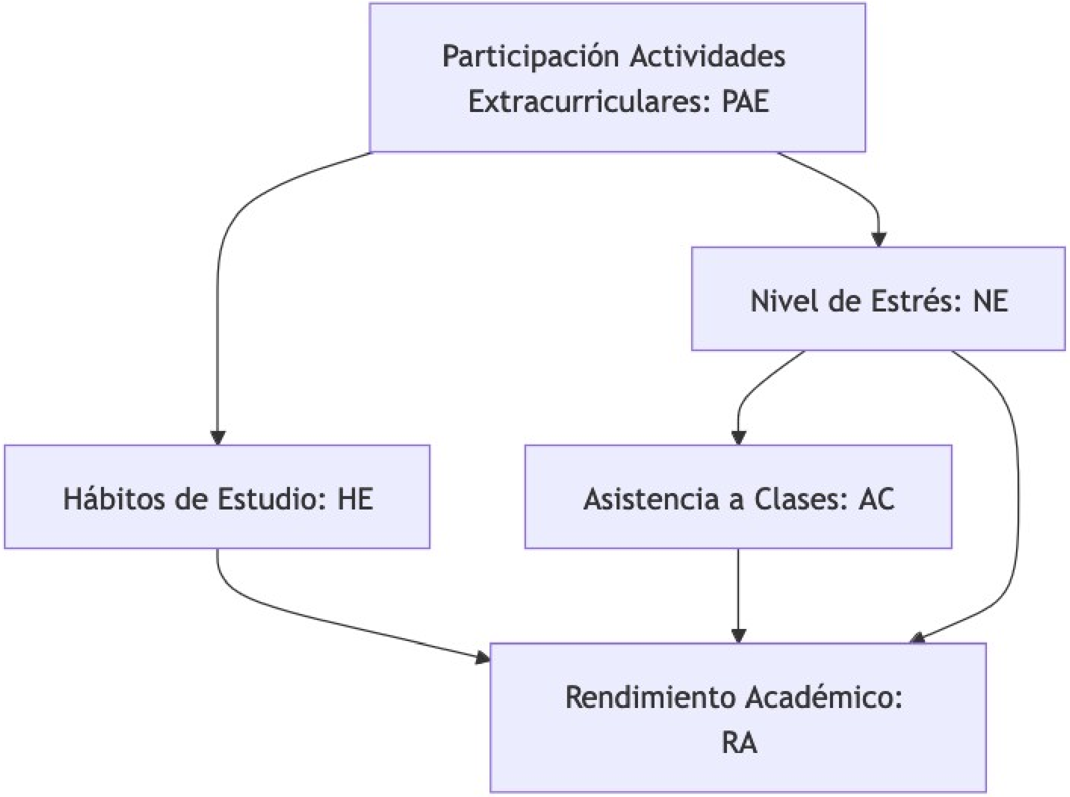


- ¿Qué distribución de prob. representa?

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Ejercicio: diseña una red bayesi.

- Solución:



$$P(\mathrm{PAE},\mathrm{HE},\mathrm{NE},\mathrm{AC},\mathrm{RA}) = P(\mathrm{PAE}) \cdot P(\mathrm{HE}|\mathrm{PAE}) \cdot P(\mathrm{NE}|\mathrm{PAE}) \cdot P(\mathrm{AC}|\mathrm{NE}) \cdot P(\mathrm{RA}|\mathrm{HE},\mathrm{AC},\mathrm{NE})$$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Ejercicio: diseña una red bayesi.

- Calcular la probabilidad de que un estudiante tenga un alto rendimiento académico dado que tiene buenos hábitos de estudio y baja asistencia a clases.
- Solución:

$$P(\mathrm{RA}|\mathrm{HE},\mathrm{AC}) = \alpha \cdot P(\mathrm{HE},\mathrm{AC}) \cdot$$

$$\sum_{\mathrm{PAE}} \left(P(\mathrm{PAE}) \cdot \sum_{\mathrm{NE}} (P(\mathrm{NE}|\mathrm{PAE}) \cdot P(\mathrm{HE}|\mathrm{PAE}) \cdot P(\mathrm{AC}|\mathrm{NE}) \cdot P(\mathrm{RA}|\mathrm{HE},\mathrm{AC},\mathrm{NE})) \right)$$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

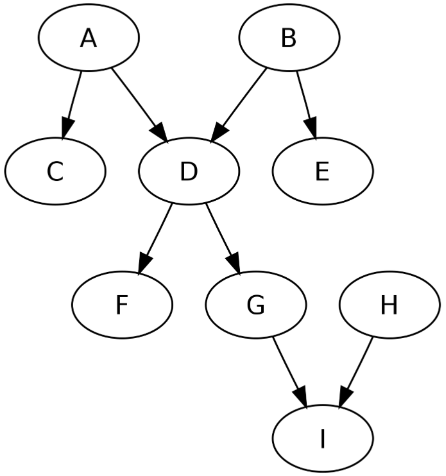
Ejercicio: diseña una red bayesi.

- Determinar qué factor tiene mayor impacto en el rendimiento académico según el modelo creado.
- Solución:
  - Mediante la estructura de la red:
    - Observando la estructura de la red, vemos que RA tiene tres padres directos: HE (Hábitos de Estudio), AC (Asistencia a Clases) y NE (Nivel de Estrés)
    - Sin valores numéricos específicos, podemos hacer algunas observaciones basadas en la estructura de la red:
      - HE, AC y NE probablemente tengan los mayores impactos directos en RA, ya que son sus padres inmediatos en la red.
      - Entre estos tres, NE podría tener un impacto ligeramente mayor debido a su influencia adicional a través de AC.
      - PAE, aunque no influye directamente en RA, tiene un impacto indirecto significativo a través de su influencia en HE y NE.
  - Información mutua (ganancia de la información en los árboles de decisión):
      - $I(X|Y) = E(X) - E(X|Y) = E(Y) – E(Y|X)$
      - Para cuantificar el impacto, podríamos calcular la Información Mutua (IM) entre cada variable y RA. La IM mide cuánta información proporciona una variable sobre otra. La variable con mayor IM con RA tendría el mayor impacto.

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Inferencia exacta en poliárboles

- Existen algoritmos más eficientes para tipos específicos de redes
  - Modelo de Kim y Pearl
    - Método de inferencia para redes bayesianas.
    - Solo aplicable a un poliárbol.
    - No existe más de un camino entre cada pareja de nodos
    - Se basa en el paso de dos tipos de mensajes entre nodos
      - Para actualizar la credibilidad
      - Para introducir nueva evidencia
  - Se puede calcular en tiempo lineal



Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Inferencia aproximada

- Sobre la inferencia exacta
  - Redes con conexión múltiple son intratables utilizando inferencia exacta
  - Complejidad NP-hard en el caso general
- Inferencia utilizando algoritmos de muestreo aleatorio (Monte Carlo)
  - Existen varios algoritmos
    - Muestreo directo
    - Muestreo por rechazo

Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Inferencia aproximada

- Muestreo directo

Algoritmo MuestreoDirecto(rb)

Entrada: rb: red Bayesiana
Salida: Un evento extraído de rb
$X \leftarrow \langle$vector de sucesos con n elementos$\rangle$
for cada variable X_i en $X_1, \ldots, X_n$ **do**
    $X_i \leftarrow$ Obtener una muestra aleatoria de $P(X_i|\text{Padres}(X_i))$
Devolver X

Para responder cualquier pregunta de la red
- Obtener un vector de eventos X[]
- Contar apariciones en X[] de las evidencias
- Dividir por suficientesMuestras

Universitat d'Alacant
Universidad de Alicante

Inferencia aproximada

Ejemplo de muestreo de una red mediante muestreo directo

P(T) = 0,001;P(R) = 0,002

P(A|T,R)=

T	R	A
0	0	0,001
0	1	0,29
1	0	0,94
1	1	0,95

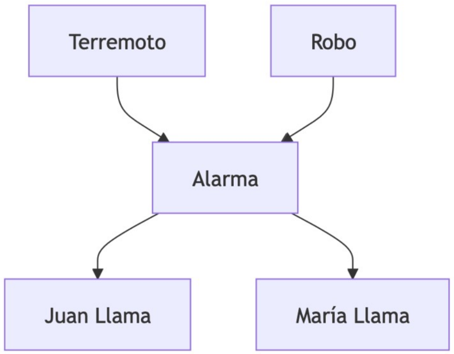
P(J|A) =

A	J
0	0,05
1	0,9

P(M|A)=

A	J
0	0,01
1	0,7

1. Muestreo a partir de P(Terremoto) = <0,001 0,999>. **supongamos (s.)** que devuelve *falso*
2. Muestreo(P(Robo)) s. devuelve *falso*
3. Muestreo(P(Alarma|<Robo=falso, Terremoto=falso>)) s. devuelve *cierto*
4. Muestreo(P(Juan|<A=cierto>)) s. Devuelve *cierto*
5. Muestreo(P(Maria|<A=cierto>)) s. Devuelve *falso*
6. *X_i = <falso,falso,cierto,cierto,falso>*



Inferencia aproximada

Ejemplo de uso

¿P(R|J,M)?

P(T) = 0,001;

P(R) = 0,002

P(A|T,R)=

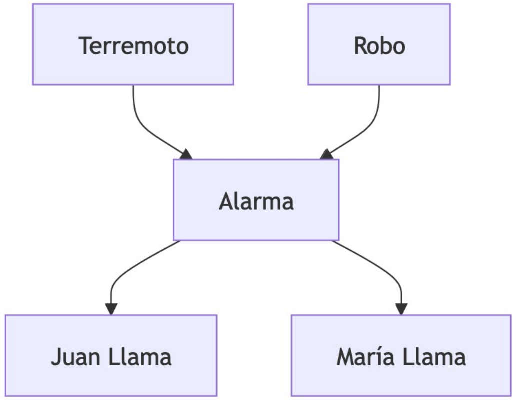
T	R	A
0	0	0,001
0	1	0,29
1	0	0,94
1	1	0,95

P(J|A) =

A	J
0	0,05
1	0,9

P(M|A)=

A	J
0	0,01
1	0,7



1. Para obtener P(R|J,M)
2. C = Contar X[ ] que cumpla este patrón X=<?,cierto,?,cierto,cierto>
3. Devolver C/numeroDeMuestras

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Inferencia aproximada

- ¿Problemas del muestreo directo?
- Otros tipos de muestreo aleatorio
  - Muestreo por rechazo
  - Gibbs sampling

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Muestreo por rechazo

Algoritmo MuestreoPorRechazo(B,c,rb)

Entradas:

B: variable buscada (pregunta)

c: valores observados de las variables conocidas C

rb: red bayesiana

Variables locales:

N: vector de recuento para cada valor de B, inicialmente 0

for $j = 1$ hasta numMuestras **do**

$x \leftarrow$ MuestreoDirecto(rb)

if x es consistente con la evidencia c **then**

$N[y] \leftarrow N[y] + 1$, donde y es el valor de B en x

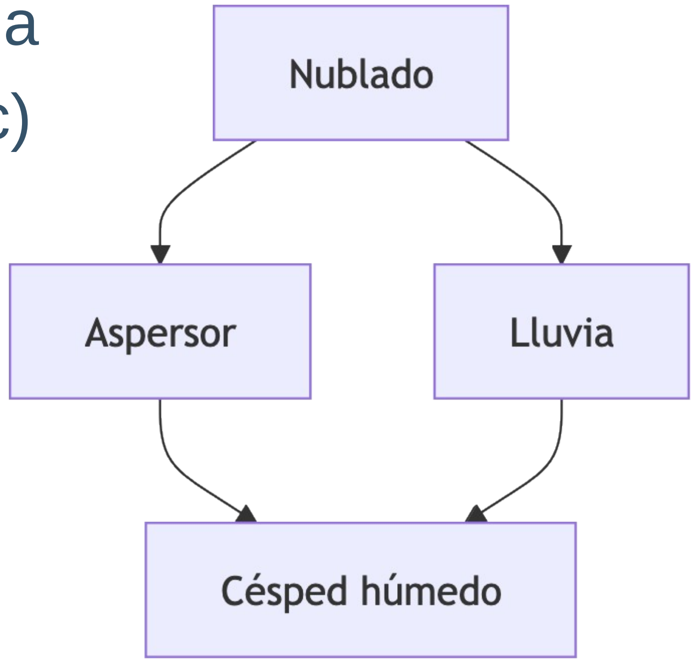
Devolver Normalizar(N)

Universitat d'Alacant
Universidad de Alicante

Muestreo por rechazo

- Ejemplo:
  - Queremos estimar P(Lluvia|Aspersor=cierto)
  - Extraemos 100 muestras. Este proceso se haría de manera incremental por muestra:
    - Si 73 tienen el aspersor apagado → se ignoran.
    - Si 27 que coinciden con la evidencia

      (x es consistente con la evidencia c)
      - En 8 Lluvia = cierto (N[cierto]++)
      - En 19 Lluvia es falso (N[falso]++)
  - P(Lluvia|Aspersor=cierto) = Normalizar(<8,19>) = <0.296,0.704>



Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Bibliografía

- Árboles:
  - Escolano et al. Inteligencia Artificial. Thomson-Paraninfo 2003. Capítulo 4.
  - Mitchel, Machine Learning. McGraw Hill, Computer Science Series. 1997
  - Cover, Thomas, Information Theory. Wiley & Sons, New York 1991
- Redes Bayesianas:
  - Stuart Russell, Peter Noving. “Inteligencia Artificial. Un enfoque Moderno” Ed. Pearson. Prentice Hall.