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

TEMA 5: Arboles de Decisión y Redes Bayesianas

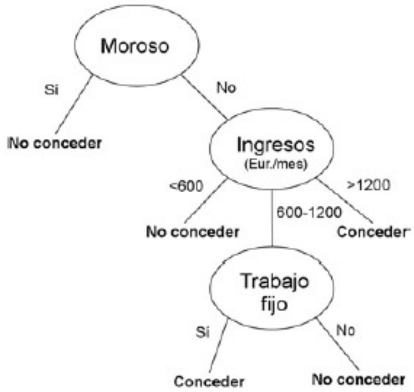
Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Parte 1: Árboles de decisión

- Árboles de decisión
  - Planteamiento del problema
  - Ejemplo: Concesión de créditos
  - Entropía y Ganancia de Información
- Algoritmo ID3
  - Algoritmo recursivo
  - Aplicación al ejemplo
  - Consideración de atributos numéricos
  - Atributos con un gran número de valores

Árboles de decisión

- Características:
  - Estructura para clasificación de vectores de atributos.
  - Establece en qué orden testar los atributos para conseguir la clasificación del vector de entrada.
  - Para componer dicho orden se eligen primero aquellos atributos que mejor ganancia de información prometen a efectos de descubrir la clase del vector de entrada.
  - Es interesante aprenderlos a partir de un conjunto de vectores



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

Ejemplo "Concesión de créditos"

Cliente	Moroso	Antigüedad (años)	Ingresos (Eur./mes)	Trab.fijo	Conceder
1	sí	>5	600-1200	sí	no
2	no	<1	600-1200	sí	sí
3	sí	1-5	>1200	si	no
4	no	>5	>1200	no	sí
5	no	<1	>1200	sí	sí
6	sí	1-5	600-1200	si	no
7	no	1-5	>1200	sí	sí
8	no	<1	<600	sí	no
9	no	>5	600-1200	no	no
10	~~no~~ sí	1-5	<600	no	no

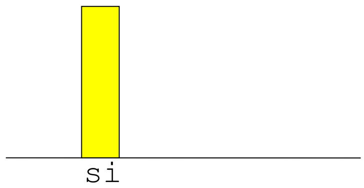
- Aprendizaje:
  - ¿Por qué atributo comenzar primero?
  - Esquema voraz: Elegir uno y filtrar recursivamente.

Entropía

- Definición:
  - Medida del grado de incertidumbre asociado a una distribución de probabilidad.
  - En una distribución uniforme, todos los valores son igualmente probables Pi = 1/N y por tanto la entropía es máxima, lo cual indica máxima incertidumbre.
  - Por el contrario, en una distribución pico en la que Pi = 1 y Pj=0, para todo j≠i la entropía es mínima lo cual indica mínima incertidumbre o sea máxima información.

$$E(S) = \sum_{i \in C} -p_i \log_2 p_i$$

-0.5log2(0.5) – 0.5log2(0.5) = 1



-1.0log2(1.0) – 0.0log2(0.0) = 0

Entropía condicionada

- Definición:
  - Entropía de la distribución de Y condicionada a X.
  - Una entropía condicionada menor que E(Y) indica que el conocimiento de X mejora la información que se dispone sobre Y

X	Y
Math	Yes
Hist.	No
CS	Yes
Math	No
Math	No
CS	Yes
Hist.	No
Math	Yes

$E(Y \mid X) = \Sigma_j \text{Prob}(X = v_j)\, E(Y \mid X = v_j)$

v_j	$\text{Prob}(X = v_j)$	$E(Y \mid X = v_j)$
Math	0.5	1
History	0.25	0
CS	0.25	0

E(Y) = 1

E(Y|X) = 0.5*1 + 0.25*0 + 0.25*0 → $\boxed{E(Y|X) = 0.5}$

Ganancia de información

- Definición:
  - Medida de cuanto ayuda el conocer el valor de una variable aleatoria X para conocer el verdadero valor de otra Y.
  - En nuestro caso, X es un atributo de un ejemplo dado mientras que Y es la clase a la que pertenece el ejemplo.
  - Una alta ganancia implica que el atributo X permite reducir la incertidumbre de la clasificación del ejemplo de entrada.

$$IG(Y \mid X) = E(Y) - E(Y \mid X)$$

X	Y
Math	Yes
History	No
CS	Yes
Math	No
Math	No
CS	Yes
History	No
Math	Yes

E(Y) = 1

E(Y|X) = 0.5

IG(Y | X) = 1 – 0.5 = 0.5

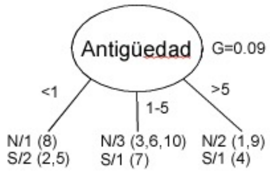
Algoritmo recursivo

Algoritmo ID3($ejemplos, atributos$) {
  Si $atributos = \emptyset$ **o** MISMACLASE($ejemplos$) {
    $C \leftarrow$ CLASEMAYORITARIA($ejemplos$)
    $N \leftarrow$ CREARNODOHOJA(C)
  }
  Sino {
    $a_{max} \leftarrow \text{máx}_{\forall A \in atributos}\, G(ejemplos, A)$
    $N \leftarrow$ CREARNODO(a_{max})
    Para cada $v_i \in$ VALORES(a_{max}) {
      $ejemplos_{v_i} \leftarrow$ {elementos de ejemplos con valor v_i para a_{max}}
      AÑADIRHIJO(N, ID3($ejemplos_{v_i}, atributos - a_{max}$))
    }
  }
  Devolver N
}

Aplicación al ejemplo

- Entropía inicial:
  - Aplicando la ecuación de entropía a los datos de entrada del ejemplo tenemos:

    E(S)= -0.4log2(0.4)-0.6log2(0.6)= 0.971



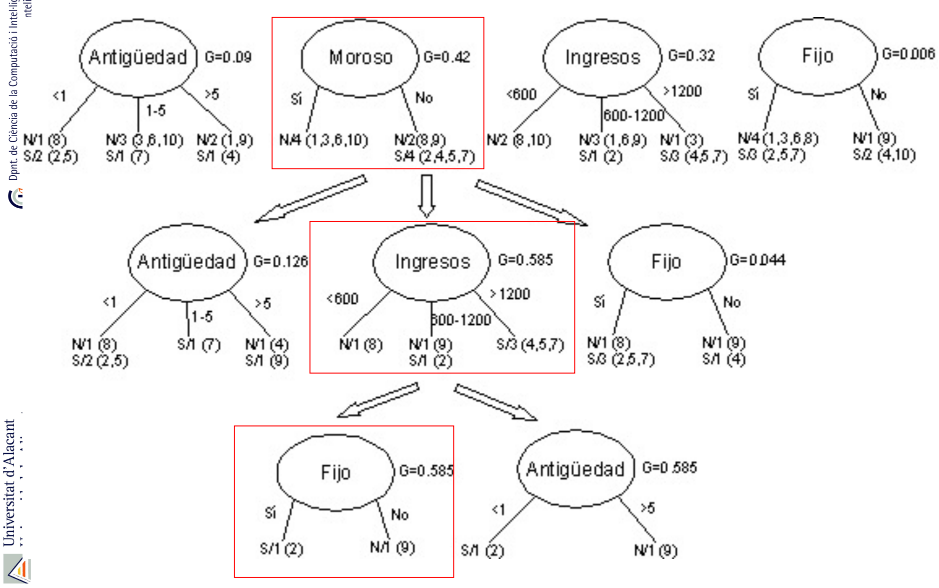
  - Para cada atributo (Antigüedad, Moroso, Ingresos, Fijo), calculamos la ganancia de información que obtenemos al seleccionar cada uno de ellos

```
Prob(S<1)=0.3,Prob(S1-5)=0.4,Prob(S>5)=0.3

E(S<1)  = -2/3log2(2/3)-1/3log2(1/3)= 0.9183
E(S1-5)= -1/4log2(1/4)-3/4log2(3/4)= 0.811
E(S>5)  = -1/3log2(1/3)-2/3log2(2/3)= 0.9183

    E(S<1)*0.3 = 0.2755
    E(S1-5)*0.4 = 0.3244
    E(S>5)*0.3 = 0.2755
H(Conceder | Antigüedad) =
    0.2755 + 0.3244 + 0.2755 = 0.8754
Ganancia = 0.971 - 0.8754 = 0.09
```


Aplicación al ejemplo



Extensiones del algoritmo

- Extensiones:
  - Atributos numéricos: ID3 sólo trabaja con atributos discretos. Si se usan atributos continuos hay que descomponerlos en rangos. Para ello se ordenan los ejemplos según el valor y se toman como puntos límite los puntos medios de aquellos en que se cambie de clase.

825 950 1150

Ejemplo	8	10	6	2	1	9	3	5	4	7
Ingresos	450	530	650	800	850	1050	1250	1400	1600	3000
Crédito	no	no	no	no	sí	no	sí	sí	sí	sí

  - Atributos con gran número de valores. Se forman grupos pequeños de ejemplos que pueden ser homogéneos por casualidad. Debe introducirse un elemento corrector que penalice atributos con un elevado número de valores (ganancia normalizada):

$$G_N(S,A) = \frac{G(S,A)}{\sum_{v_i \in V(A)} - p_{v_i} \log_2 p_{v_i}}$$

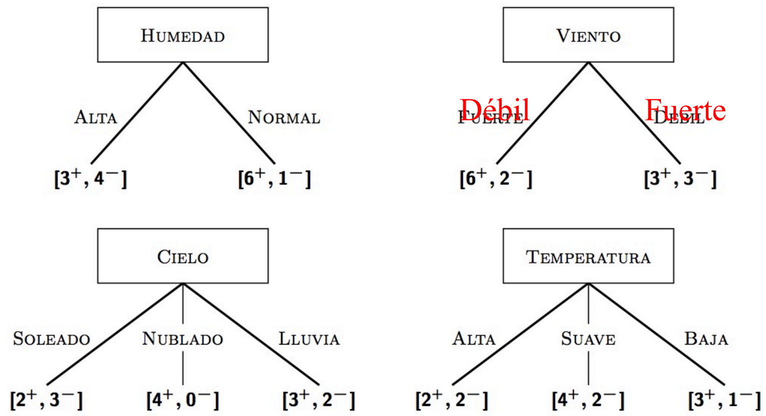
  - Sobre-entrenamiento. Comprobación de capacidad de generalización.

Ejercicios

Objetivo: Dado el conjunto de entrenamiento, aprender el concepto “Días en los que se juega al tenis” obteniendo el árbol de decisión mediante el algoritmo ID3

Ej.	Cielo	Temperatura	Humedad	Viento	JugarTenis
D_1	Soleado	Alta	Alta	Débil	-
D_2	Soleado	Alta	Alta	Fuerte	-
D_3	Nublado	Alta	Alta	Débil	+
D_4	Lluvia	Suave	Alta	Débil	+
D_5	Lluvia	Baja	Normal	Débil	+
D_6	Lluvia	Baja	Normal	Fuerte	-
D_7	Nublado	Baja	Normal	Fuerte	+
D_8	Soleado	Suave	Alta	Débil	-
D_9	Soleado	Baja	Normal	Débil	+
D_{10}	Lluvia	Suave	Normal	Débil	+
D_{11}	Soleado	Suave	Normal	Fuerte	+
D_{12}	Nublado	Suave	Alta	Fuerte	+
D_{13}	Nublado	Alta	Normal	Débil	+
D_{14}	Lluvia	Suave	Alta	Fuerte	-

Ejercicios



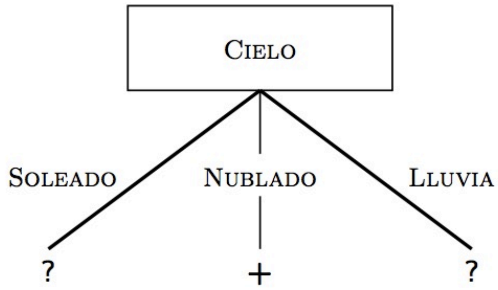
Ejercicios

- Entropía inicial: $\mathbf{Ent([9^+, 5^-]) = 0{,}94}$
- Selección del atributo para el nodo raíz:
  - $\mathbf{Ganancia(D,\textsc{Humedad}) =}$
    $\mathbf{0{,}94 - \frac{7}{14} \cdot Ent([3^+, 4^-]) - \frac{7}{14} \cdot Ent([6^+, 1^-]) = 0{,}151}$
  - $\mathbf{Ganancia(D,\textsc{Viento}) =}$
    $\mathbf{0{,}94 - \frac{8}{14} \cdot Ent([6^+, 2^-]) - \frac{6}{14} \cdot Ent([3^+, 3^-]) = 0{,}048}$
  - $\mathbf{Ganancia(D,\textsc{Cielo}) =}$
    $\mathbf{0{,}94 - \frac{5}{14} \cdot Ent([2^+, 3^-]) - \frac{4}{14} \cdot Ent([4^+, 0^-])}$
    $\mathbf{- \frac{5}{14} \cdot Ent([3^+, 2^-]) = 0{,}246}$ (mejor atributo)
  - $\mathbf{Ganancia(D,\textsc{Temperatura}) =}$
    $\mathbf{0{,}94 - \frac{4}{14} \cdot Ent([2^+, 2^-]) - \frac{6}{14} \cdot Ent([4^+, 2^-])}$
    $\mathbf{- \frac{4}{14} \cdot Ent([3^+, 1^-]) = 0{,}02}$
- El atributo seleccionado es CIELO

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

Ejercicios

Árbol parcialmente construido:



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

Ejercicios

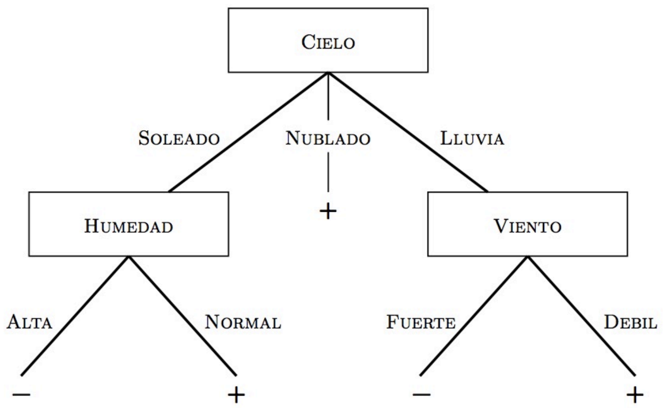
- Selección del atributo para el nodo CIELO=SOLEADO
- $\mathbf{D}_{\text{SOLEADO}} = \{\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_8, \mathbf{D}_9, \mathbf{D}_{11}\}$ con entropía $\mathbf{Ent([2^+, 3^-]) = 0{,}971}$
  - $\mathbf{Ganancia(D_{\text{SOLEADO}}, \text{HUMEDAD}) =}$ $\mathbf{0{,}971 - \frac{3}{5} \cdot 0 - \frac{2}{5} \cdot 0 = 0{,}971}$ (mejor atributo)
  - $\mathbf{Ganancia(D_{\text{SOLEADO}}, \text{TEMPERATURA}) =}$ $\mathbf{0{,}971 - \frac{2}{5} \cdot 0 - \frac{2}{5} \cdot 1 - \frac{1}{5} \cdot 0 = 0{,}570}$
  - $\mathbf{Ganancia(D_{\text{SOLEADO}}, \text{VIENTO}) =}$ $\mathbf{0{,}971 - \frac{2}{5} \cdot 1 - \frac{3}{5} \cdot 0{,}918 = 0{,}019}$
- El atributo seleccionado es HUMEDAD

Ejercicios

- Selección del atributo para el nodo CIELO=LLUVIA:
- $D_{\text{LLUVIA}} = \{D_4, D_5, D_6, D_{10}, D_{14}\}$ con entropía $\text{Ent}([3^+, 2^-]) = 0{,}971$
  - $\text{Ganancia}(D_{\text{LLUVIA}}, \text{HUMEDAD}) = 0{,}971 - \frac{2}{5} \cdot 1 - \frac{3}{5} \cdot 0{,}918 = 0{,}820$
  - $\text{Ganancia}(D_{\text{LLUVIA}}, \text{TEMPERATURA}) = 0{,}971 - \frac{3}{5} \cdot 0{,}918 - \frac{2}{5} \cdot 1 = 0{,}820$
  - $\text{Ganancia}(D_{\text{LLUVIA}}, \text{VIENTO}) = 0{,}971 - \frac{3}{5} \cdot 0 - \frac{2}{5} \cdot 0 = 0{,}971$ (mejor atributo)
- El atributo seleccionado es VIENTO

Ejercicios

- Árbol finalmente aprendido:



Parte 2: Redes Bayesinas

Probabilidad como medida de incertidumbre

Teorema de Bayes

Redes Bayesianas

Inferencia mediante redes Bayesianas

- Inferencia Exacta
- Ejemplos
- Inferencia aproximada
  - Muestreo directo
  - Muestreo por rechazo
  - Muestreo Gibbs

Para saber más

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

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

Constante de normalización P(*B*)

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

Regla de la cadena

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

Redes Bayesianas (I)

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

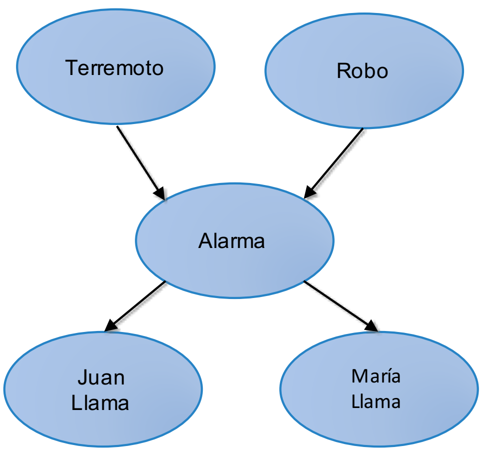
Su finalidad principal es calcular la distribución conjunta de las variables nodo

21

Redes Bayesianas (II) Semántica

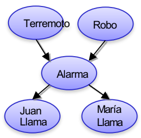
Dada la siguiente red bayesiana,

¿Qué distribución representa



Redes Bayesianas (III)
Semántica

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

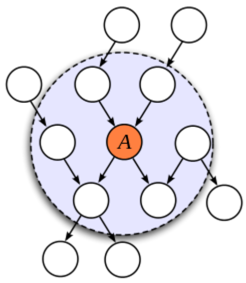


P(T) · P(R) · P(A|T,R) · P(J|A) · P(M|A)

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

Redes Bayesianas (IV)
Semántica

- 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



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

Sistemas Inteligentes

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 (I)

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)

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

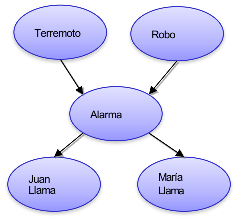
Problema: Mucha complejidad

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

Inferencia exacta(II)
Ejemplo 1

- ¿Cuál es la probabilidad de 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)

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 (III)
Ejemplo 1

De esta manera tenemos que:

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

Distribución de probabilidad

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

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

Inferencia (IV)

Ejemplo 2

¿P(R|J+,M+)? Si sabemos que:

P(T) = 0,001

P(R) = 0,002

P(J|A) =

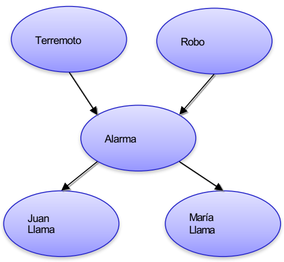
A	J
0	0,05
1	0,9

P(A|T,R)=

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

P(M|A)=

A	M
0	0,01
1	0,7



Inferencia (V)

Ejemplo 2

De esta manera tenemos que:

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

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

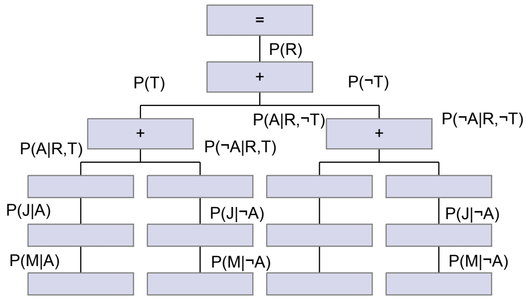
$$\alpha \cdot P(R) \cdot \sum_T \left(P(T) \cdot \sum_A \left(P(A \mid R,T) \cdot P(J \mid A) \cdot P(M \mid A) \right) \right)$$

Inferencia (VI)

Ejemplo 2

Para calcular descomponemos utilizando un árbol:

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



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

Inferencia (VII)

Ejemplo 2

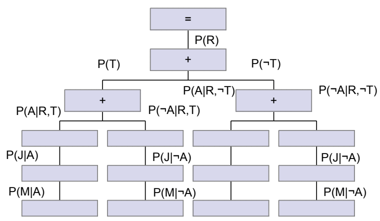
P(R|J,M)

1	P(A|R,T)*P(J|A)*P(M|A)	0,5985
	P(¬A|R,T)*P(J,¬A)*P(M|¬A)	0,000025
	P(T)* SUM 1	0,00059853
2	P(A|R,¬T)*P(J|A)*P(M|A)	0,5922
	P(¬A|R,¬T)*P(J,¬A)*P(M|¬A)	0,00003
	P(¬T) * SUM 2	0,59163777
	TOTAL R+	**0,00118447**

P(¬R|J,M)

1	P(A|¬R,T)*P(J|A)*P(M|A)	0,1827
	P(¬A|¬R,T)*P(J,¬A)*P(M|¬A)	0,000355
	P(T)* SUM 1	0,00018306
2	P(A|¬R,¬T)*P(J|A)*P(M|A)	0,00063
	P(¬A|¬R,¬T)*P(J,¬A)*P(M|¬A)	0,0004995
	P(¬T) * SUM 2	0,00112856
	TOTAL ¬R	**0,00130899**

Cómo alpha*(R,¬R) = 1



P(R|J,M)=	**R**	**0,47503178**
	¬R	**0,52496822**

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 (VIII)

Ejercicio 1

¿P(J|R)?

$$P(J \mid R) = \sum_{T}\sum_{A}\sum_{M} P(R)\cdot P(T)\cdot P(A \mid R,T)\cdot P(J \mid A)\cdot P(M \mid A) =$$

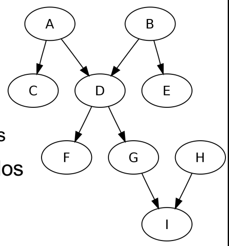
$$= P(R)\cdot \sum_{T}(P(T)\cdot \sum_{A}(P(A) \mid R,T)\cdot P(J \mid A)\cdot \sum_{M} P(M \mid A)))$$

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



Inferencia aproximada (I)

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
  - Gibbs Sampling

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 (II)
Muestreo directo

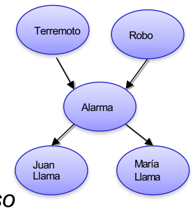
- rb: red Bayesiana
- ALGORITMO **Muestreo_Directo**(*rb*) retorna un evento extraido de *rb*
  - X = <vector de sucesos con n elementos>
  - Para cada variable X_i en $X_1,\ldots,X_n$ hacer
    - X_i = Obtener una muestra aleatoria de $P(X_i|Padres(X_i))$
  - Devolver X

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

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 (III)

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 = <falso,falso,cierto,cierto,falso>*

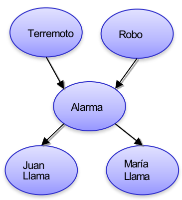
Sistemas Inteligentes

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 (IV)

Ejemplo de uso

¿P(R|J,M)?



P(T) = 0,001;

P(R) = 0,00

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

Inferencia aproximada (V)

¿Problema del muestreo directo?

Otros tipos de muestreo aleatorio

- Muestreo por rechazo
- Gibbs Sampling

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 (VI)
Muestreo por rechazo

- ALGORITMO **Muestreo_por_Rechazo**(*B,c,rb*) retorna estimación P(B|c)
- Para j = 1 hasta num_muestras hacer
  - **x** = Muestro_Directo(rb)
  - Si **x** es consistente con la evidencia c:
  - N[y] = N[y] +1, donde y es el valor de B en **x**
- Devolver Normalizar(N)

Entradas:

- B: variable buscada (pregunta)
- c: valores observados de las variables conocidas C
- rb: red bayesiana

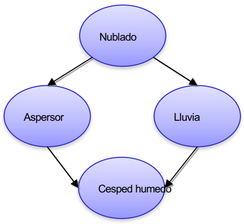
Variables locales:

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

Universitat d'Alacant
Universidad de Alicante

Inferencia aproximada (VII)
Muestreo por rechazo

Ejemplo:



Queremos estimar P(Lluvia|Aspersor=cierto)
Extraemos 100 muestras, de las cuales 73 tienen el aspersor apagado
Nos quedamos con las 27 que coinciden con la evidencia
De las 27:

- En 8 Lluvia = cierto
- En 19 Lluvia es falso

P(Lluvia|Aspersor=cierto) = Normalizar(<8,19>) = <0.296,0.704

Inferencia aproximada (VIII)
Muestreo por Gibbs (MCMC)

- ALGORITMO **Muestreo_por_Gibbs**(*B*,*c*,*rb*,*N*) retorna estimación P(B|c)
- Inicializar **x** con valor aleatorios para las variables en Z
- Para j = 1 hasta num_muestras hacer
  - Para cada Z_i en Z hacer
    - **x**[Z_i] = muestrear P(Z_i|mb(Z_i))
    - N[x] = N[x] + 1 donde x es el valor de B en **x**
- Devolver Normalizar(N)

Entradas:

- B: variable buscada (pregunta)
- c: valores observados de las variables conocidas C
- rb: red bayesiana

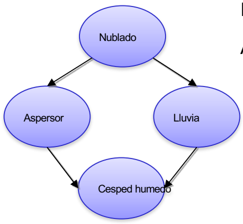
Variables locales:

- N: vector de recuento para cada variable B (inicialmente vale 0)
- Z, las variables sin evidencia en rb
- x: el estado de la red, copiado inicialmente de c

Funciones

- mb() retorna la cobertura de Markov un nodo

Inferencia aproximada (IX)
Ejemplo



Queremos estimar P(Lluvia|Aspersor=cierto,Cesped=cierto)

Las variables conocidas: Aspersor y Cesped, se fijan a su valor.

Las variables desconocidas: Nublado y Lluvia se establecen aleatoriamente.

Imaginemos que el estado inicial es: [cierto,cierto,falso,cierto]

Ahora las variables sin evidencia se muestrean repetidamente en orden arbitrario. Por ejemplo

- Se muestrea Nublado, dado su recubrimiento de Markov. Por tanto extraemos de P(Nublado|Aspersor=cierto,Lluvia=Falso) Asumamos que el resultado es falso y por tanto el nuevo estado es [falso,cierto,falso,cierto]
- Se muestrea Lluvia dado su recubrimiento: P(Lluvia|Nublado=falso, aspersor=cierto, cesped=cierto). Asumamos que el resultado es cierto. El nuevo estado es [falso,cierto,cierto,cierto]

Todo estado visitado mediante este proceso es una muestra que contribuye a estimar la pregunta

Por ejemplo, si durante es te proceso se visitan 20 estados donde lluvia es cierto y 60 donde es falso
P(L|A=cierto,C=cierto)=Normalizar(<20,60>)=<0.25,0.75>

Bibliografía

- 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