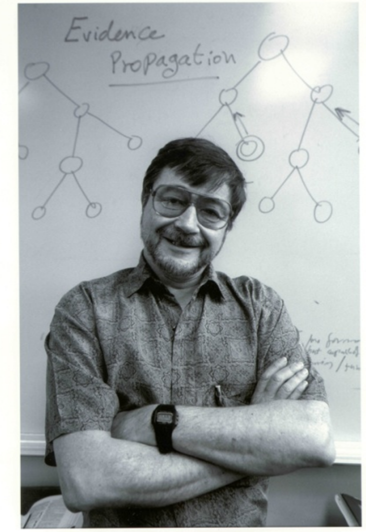
Redes bayesianas

Competencias



Judea Pearl

Al finalizar el tema, el estudiante deberá ser capaz de....

- **Modelar** una situación real
  con una red bayesiana

  Nodos, enlaces, parámetros

- Entender el significado e importancia de las condiciones de **independencia condicional**
- **Aplicar algoritmos de propagación y aprendizaje básicos**
- **Predecir la evolución de las probabilidades** en la red conforme se adquiere nueva evidencia
- **Manejar herramientas** de implementación de redes bayesianas (GENIE)

Índice

- Repaso de conceptos básicos de probabilidad
- Presentación intuitiva
- Definición formal de red bayesiana
- Teorema fundamental
- Modelado con redes bayesianas
- Algoritmo de propagación de probabilidades en árboles

Repaso de probabilidad (I)

- Definición de **probabilidad condicionada**:

  $P(X/Y) = P(X,Y)/P(Y)$

- Dos variables X e Y son **independientes** si se tiene que

  $P(X/Y) = P(X)$

  Caracterización de la independencia

  X e Y son independientes sí y sólo sí $P(X,Y) = P(X) \cdot P(Y)$

- Dos variables X e Y son **condicionalmente independientes** dado una tercera variable Z si se tiene que

  $P(X/Y,Z) = P(X/Z)$

  Caracterización de la independencia condicional:

  X e Y son independientes dado Z sí y sólo sí $P(X,Y/Z) = P(X/Z)P(Y/Z)$

  También se dice que **Z separa condicionalmente a X e Y**

Repaso de probabilidad (II)

- **Teorema de Bayes (teorema de inversión)**

$$P(X/Y) = \frac{P(Y/X)P(X)}{P(Y)}$$

- **Ley de probabilidad total**

Sean $Y_1, \ldots, Y_n$ un conjunto de variables exhaustivo y excluyente. Entonces:

$$P(X) = \sum_{i=1,\ldots,n} P(X/Y_i)P(Y_i)$$

- **Generalización de la ley de probabilidad total**

$$P(Y_1/Y_2) = \sum P(Y_1/X, Y_2)\, P(X/Y_2)$$

Repaso de probabilidad (III)

Cálculo de distribución marginal a partir de la distribución conjunta

Dado un conjunto de variables aleatorias $X_1,...,X_n$, y sea $p(X_1,...,X_n)$ su distribución de probabilidad conjunta. La distribución marginal de X_j se calcula mediante:

$$P(X_j=x_j) = \sum_{X_1,...,X_j=x_j,...X_n} P(X_1,...,X_n)$$

Método probabilístico clásico

Individuos	Edad	Obesidad	Hernia	Indigestión	Vómitos
Individuo 1	Mayor_50	no	no	no	no
Individuo 2	Mayor_50	no	no	no	no
Individuo 3	Mayor_50	no	no	no	no
Individuo 4	Mayor_50	no	no	no	no
Individuo 5	Mayor_50	no	sí	no	sí
Individuo 6	Mayor_50	sí	sí	no	sí
Individuo 7	Mayor_50	sí	sí	no	sí
Individuo 8	Mayor_50	sí	sí	no	si
Individuo 9	Menor_50	no	no	no	no
Individuo 10	Menor_50	no	no	no	no
Individuo 11	Menor_50	no	no	no	no
Individuo 12	Menor_50	no	no	no	no
Individuo 13	Menor_50	no	no	no	no
Individuo 14	Menor_50	no	no	no	no
Individuo 15	Menor_50	no	no	no	no
Individuo 16	Menor_50	no	no	no	no
Individuo 17	Menor_50	no	no	sí	sí
Individuo 18	Menor_50	sí	no	no	sí
Individuo 19	Menor_50	sí	no	sí	sí
Individuo 20	Menor_50	sí	no	no	no

Método probabilístico clásico

Se pide,

a) Calcular la distribución conjunta de las variables

b) A partir de la distribución conjunta calcula

- Las distribuciones marginales P(Edad = Mayor_50) y P(Edad = Mayor_50, Hernia= sí)
- P(Hernia=si/Vómitos=si) (diagnóstico)
- P(Vómitos=si/Obesidad=si, Hernia=si) (predicción)

Redes bayesianas

Índice

Presentación intuitiva

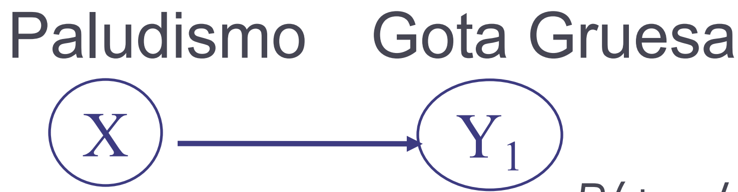
En una red bayesiana, cada nodo corresponde a una variable, que a su vez representa una entidad del mundo real

Notación:

- *X,* variable o nodo
- *x*, valor (*+x*=sí, ¬*x*=no)

Los arcos que unen los nodos representan relaciones de *influencia causal*

Presentación intuitiva: ejemplo 1



$P(+x) = 0.003$

$P(+y_1/+x) = 0.992$
$P(+y_1/\neg x) = 0.0006$

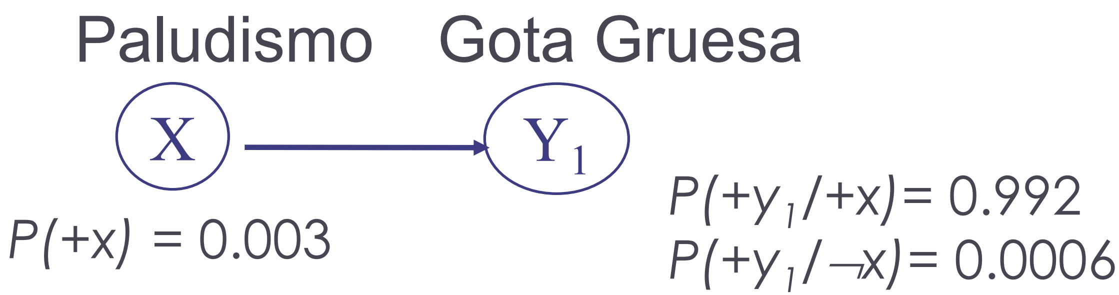
Información cualitativa

- Nodos (variables)
- Enlaces (relaciones de influencia causal)

Información cuantitativa:

- La probabilidad a priori de los nodos que no tienen padres.
- La probabilidad condicionada de los nodos con padres.

Presentación intuitiva: significado de los parámetros



Enfermedad

- $P(+x) = \mathbf{0.003}$ <u>prevalencia</u>

Test

- $P(+y_1/+x) = \mathbf{0.992}$ <u>sensibilidad</u>
- $P(\neg y_1/+x) = 0.008$ falsos negativos
- $P(+y_1/\neg x) = \mathbf{0.0006}$ falsos positivos
- $P(\neg y_1/\neg x) = 0.9994$ <u>especificidad</u>

En medicina siempre se buscan las pruebas con **mayor grado de sensibilidad y especificidad**

Presentación intuitiva: inferencias

Inferencias a partir de los datos

- La probabilidad a priori de Y_1,

$$P(+y_1) = P(+y_1/+x)\,P(+x) + P(+y_1/\neg x)\,P(\neg x) = 0.00357$$

$$P(\neg y_1) = P(\neg y_1/+x)\,P(+x) + P(\neg y_1/\neg x)\,P(\neg x) = 0.99643$$

- Las probabilidades a posteriori dada una evidencia observada ***e***, $P^*(x) = P(x/\boldsymbol{e})$.

Supongamos que el test de la gota gruesa ha dado positivo:

$$P^*(+x) = P(+x/+y_1) = \frac{P(+x)\ P(+y_1/+x)}{P(+y_1)} = \frac{0.003\cdot 0.992}{0.00357} = 0.83263$$

$$P^*(\neg x) = P(\neg x/+y_1) = \frac{P(\neg x)\ P(+y_1/\neg x)}{P(+y_1)} = \frac{0.997\cdot 0.0006}{0.00357} = 0.16737$$

Presentación intuitiva: inferencias

La expresión general del teorema de Bayes que hemos utilizado es:

$$P^*(x) = P(x/y) = \frac{P(y / x)P(x)}{P(y)}$$

Que podemos reescribir como;

$$P^*(x) = \alpha\, P(x)\, \lambda(x)$$

Donde $\alpha = [P(y)]^{-1}$ y $\lambda(x) = P(y/x)$.

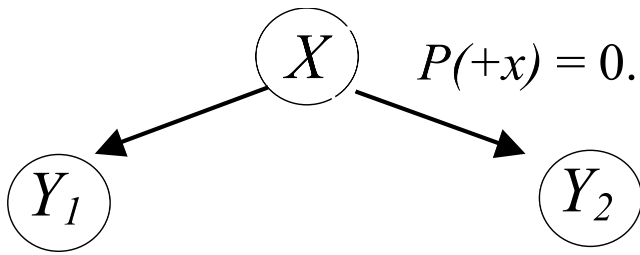
Utilizando esta nueva expresión:

$$P^*(+x) = \alpha\, 0.003\ 0.992 = 0.00298\, \alpha$$

$$P^*(\neg x) = \alpha\, 0.997\ 0.0006 = 0.000598\, \alpha$$

Y normalizando obtenemos el mismo resultado que antes.

Presentación intuitiva: ejemplo 2



$P(+x) = 0.003$

Y_2= fiebre

$P(+y_1/+x)= 0.992$

$P(+y_1/\neg x)= 0.0006$

$P(+y_2/+x)= 0.98$

$P(+y_2/\neg x)= 0.017$

Inferencias

a) $e = \{+y_2\}$

$$P^*(+x) = P(+x/+y_2) = \alpha\, 0.003\ 0.98 = 0.00294\ \alpha = 0.148$$

$$P^*(\neg x) = P(\neg x/+y_2) = \alpha\, 0.997\ 0.0017 = 0.016949\ \alpha\ = 0.852$$

Presentación intuitiva: ejemplo 2

b) $e = \{+y_1, +y_2\}$

$$P^*(x) = P(x/y_1, y_2) = \frac{P(x)\, P(y_1, y_2/x)}{P(y_1, y_2)} \; ?$$

Hipótesis independencia condicional:

$$P(y_1,y_2/x) = P(y_1/x)\, P(y_2/x)$$

$$\lambda(x) = \lambda_{Y1}(x)\, \lambda_{Y2}(x)$$

Y entonces:

$$P^*(x) = \alpha\, P(x)\, \lambda(x)$$

Presentación intuitiva: ejemplo 2

Si seguimos con los cálculos:

$P^*(+x) = 0.9663$

$P^*(\neg x) = 0.0337$

Ejercicio propuesto:

Utilizando esta formulación, hallar P*(x) si $e = \{\neg y_1, +y_2\}$

c) Sea **e** = $\{+y_2\}$. ¿Cuál será el resultado del test?

$$P^*(y_1) = P(y_1/y_2) = \sum P(y_1/x, y_2)\, P(x/y_2) = \sum P(y_1/x)\, P(x,y_2) P(y_2)^{-1}$$

Independencia condicional

Sea $\pi_{Y1}(x) = P(x, y_2) = P(x)\, P(y_2/x)$

$\alpha = [P(y_2)]^{-1}$

Presentación intuitiva: ejemplo 2

Entonces $P^*(y_1) = \alpha \sum P(y_1/x)\ \pi(x)$

Y finalmente:

$P^*(+y_1) = \alpha\ [\pi_{Y1}(+x)\ P(+y_1/+x) + \pi_{Y1}(\neg x)\ P(+y_1/\neg x)] = 0.14715$

$P^*(\neg y_1) = \alpha\ [\pi_{Y1}(+x)\ P(\neg y_1/+x) + \pi_{Y1}(\neg x)\ P(\neg y_1/\neg x)] = 0.85285$

Conclusiones

Las redes bayesianas permiten hacer inferencias:

- **Abductivas**: ¿cúal es el diagnóstico que mejor explica los hallazgos?
- **Predictivas**: ¿cuál es la probabilidad de obtener cierto resultado en el futuro?

Redes bayesianas

Índice

Definición formal de red bayesiana

Una red bayesiana es:

- Un conjunto de variables **proposicionales** V;
- Un **conjunto de relaciones binarias** definida sobre las variables de V, E:
- Una **distribución de probabilidad conjunta P** sobre las variables de V,

tales que:

- (V, E) forman un grafo **acíclico**, **conexo** y **dirigido** G.
- (G, P) cumplen las hipótesis de **independencia condicional**, también llamadas de separación direccional

Definición formal de red bayesiana

Definición de variable proposicional

Una variable proposicional es una variable aleatoria que toma un conjunto *exhaustivo* y *excluyente* de valores.

Definición: hipótesis de independencia condicional

Un grafo acíclico conexo y dirigido $G = (V, E)$ y una distribución de probabilidad conjunta P definida sobre las variables del grafo se dice que cumplen las hipótesis de independencia condicional o separación direccional, si

$$\forall X \in V \text{ y } \forall Y \in V - \{X \cup de(X) \cup pa(X)\}$$

se tiene que

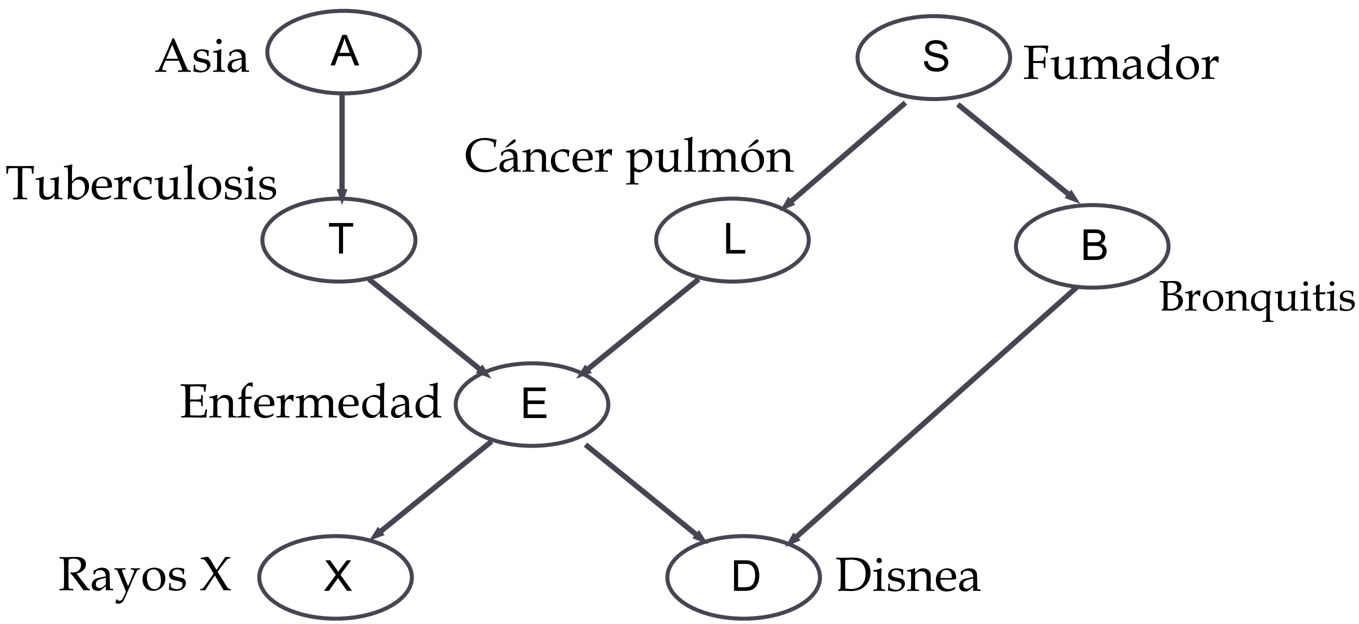
$$X \text{ es independiente de } Y \text{ dado } pa(X)$$

Definición formal de red bayesiana: ejercicios

Ejercicio 1. ¿Qué independencias implica la red?

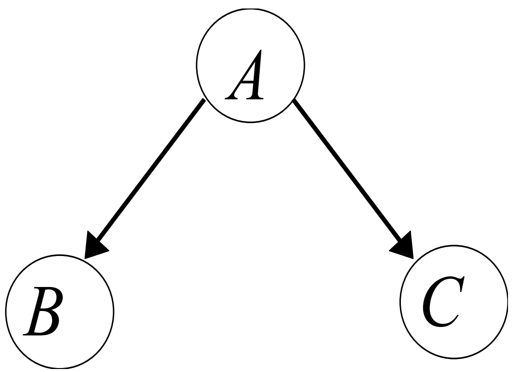
Definición formal de red bayesiana: Ejercicios

Ejercicio 2. ¿Qué independencias implica la red Asia?



Definición formal de red bayesiana: Ejercicios

Ejercicio 3. Comprobar si esta red es bayesiana



$P(a_1,b_1,c_1) = 0.084$ $P(a_1,b_1,c_2) = 0.036$

$P(a_1,b_2,c_1) = 0.126$ $P(a_1,b_2,c_2) = 0.054$

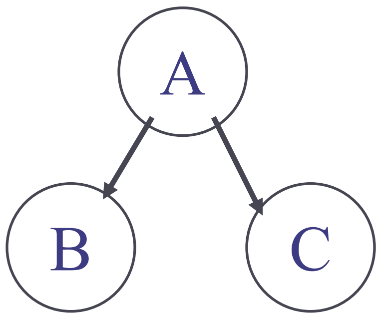
$P(a_2,b_1,c_1) = 0.084$ $P(a_2,b_2,c_1) = 0.056$

$P(a_2,b_1,c_2) = 0.336$ $P(a_2,b_2,c_2) = 0.224$

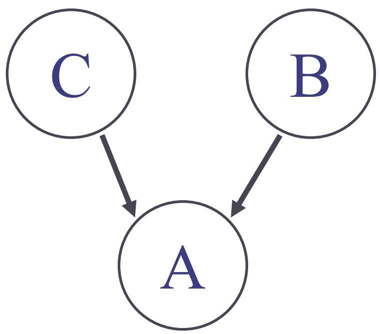
Definición formal de red bayesiana: Ejercicios

Ejercicio 4:

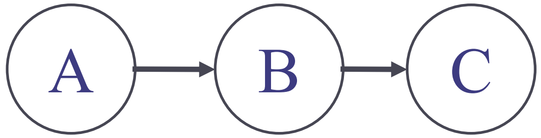
Comprobar qué relaciones de independencia condicional deben darse para todas las posibles estructuras de redes con tres nodos



Cola con cola



Cabeza con cabeza



Cabeza con cola

Índice

Teorema fundamental (Factorización de la probabilidad)

Dada una red bayesiana, su distribución de probabilidad puede expresarse como:

$$P(x_1, \ldots, x_n) = \prod_{X_i} P(x_i / pa(x_i))$$

Demostración:

Sea $\{X_1, \ldots, X_n\}$ una ordenación de las variables en la que los padres de cada nodo aparezcan siempre después de el. Entonces:

$$P(x_1, \ldots, x_n) = \prod_{X_i} P(x_i / x_{i+1}, \ldots, x_n)$$

Pero por la forma de escoger la ordenación, el conjunto $\{X_{i+1}, \ldots, X_n\}$ incluye a todos los padres de X, y, en consecuencia, la separación direccional nos dice que

$$P(x_i / x_{i+1}, \ldots, x_n) = P(x_i / pa(x_i))$$

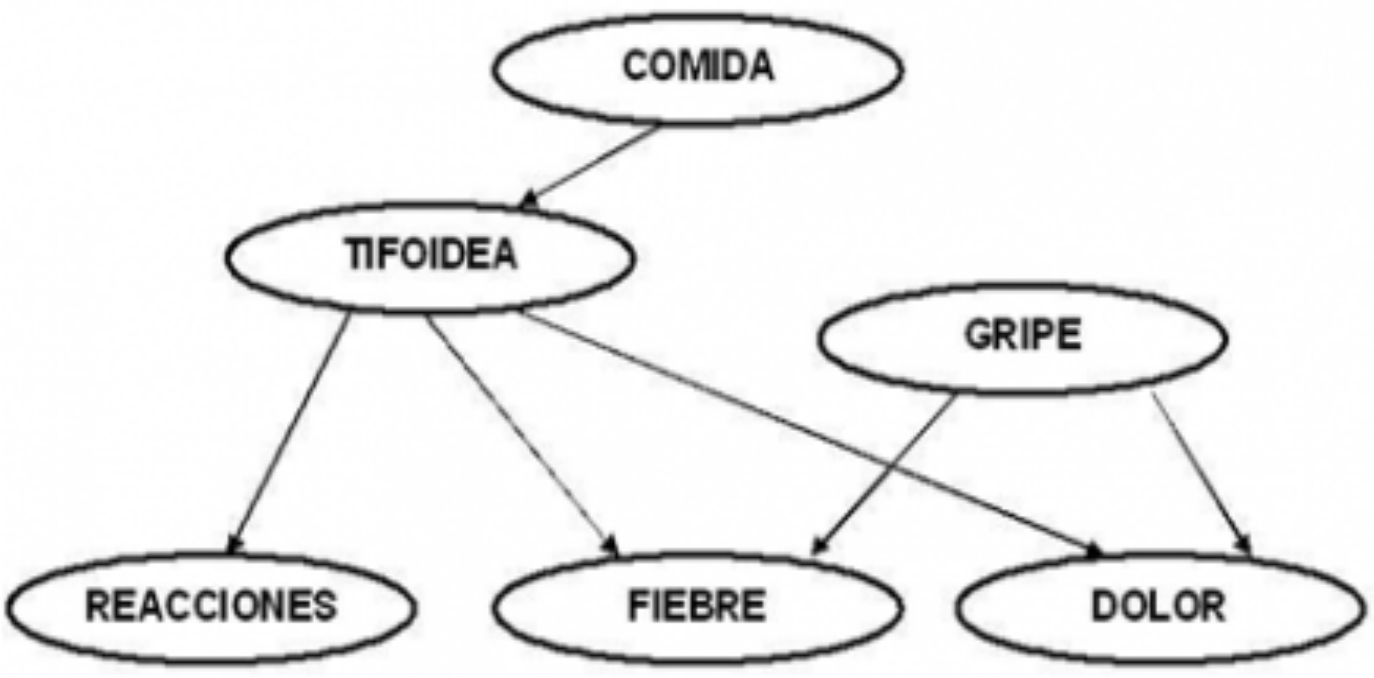
Teorema fundamental: Importancia del teorema

Nos permite describir una red bayesiana a partir de las probabilidades condicionadas de cada nodo dados sus padres en lugar de a partir de la probabilidad conjunta, que:

- requiere un ***número de parámetros exponencial*** en el número de nodos.
- plantea el problema de ***verificar la independencia condicional***

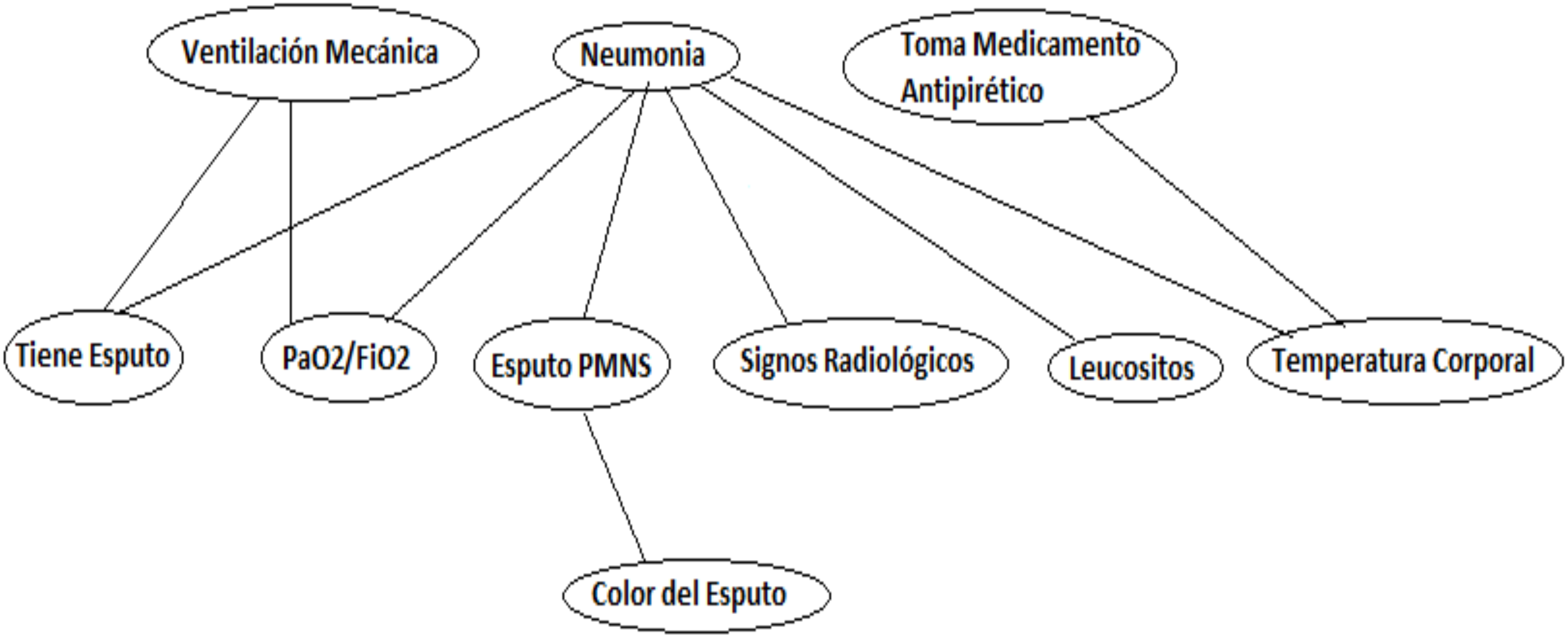
Teorema fundamental: ejercicio 1

Aplica el teorema fundamental a las siguiente red:

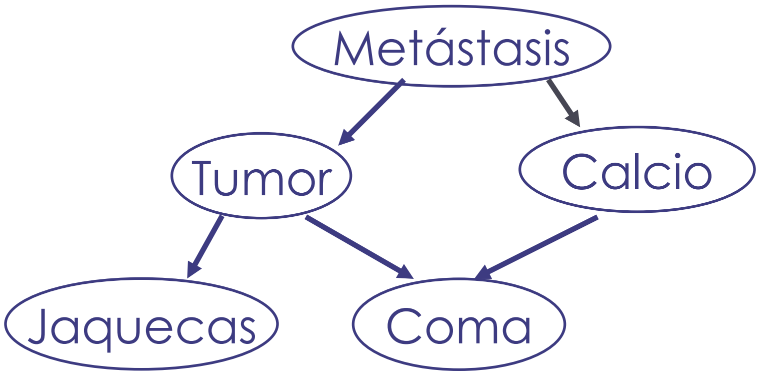


Teorema fundamental: ejercicio 2

Aplica el teorema fundamental a las siguiente red:



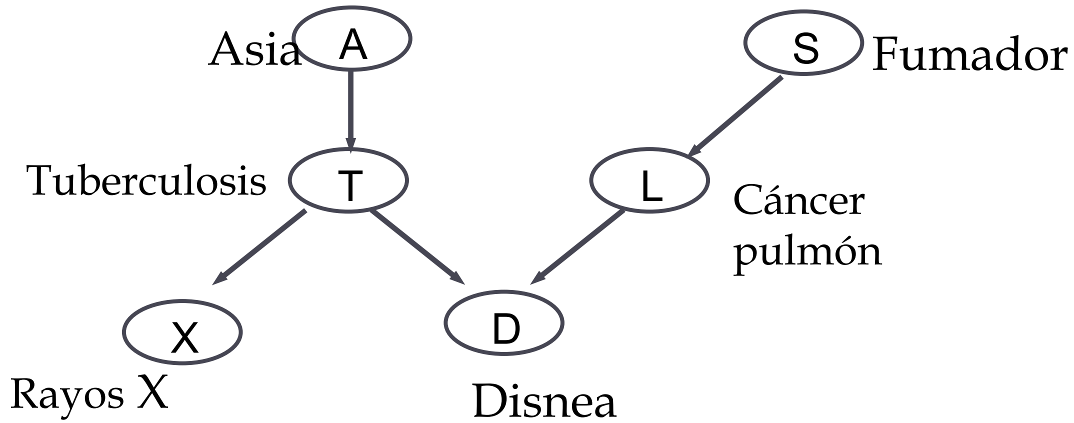
Ejercicio de repaso 1



Se pide:

- ¿Qué independencias/dependencias entre las variables de la red implican las hipótesis de independencia condicional?
- Si suponemos ciertas las hipótesis de independencia condicional, ¿cuántas probabilidades sería necesario especificar?. Dar estos valores de una forma coherente con el sentido común.
- Si no podemos suponer las hipótesis de independencia condicional, ¿qué probabilidades deberíamos pedir al experto? ¿Cuántos valores son, en total?
- ¿Cómo podemos calcular la probabilidad conjunta a partir de las condicionadas?. Aplicando el teorema de factorización, indica cómo se calcularía la probabilidad de que el paciente tenga metástasis dado que esta en coma

Ejercicio de repaso 2



Se pide:

- ¿Qué independencias/dependencias entre las variables de la red implican las hipótesis de independencia condicional?
- Si suponemos ciertas las hipótesis de independencia condicional, ¿cuántas probabilidades sería necesario especificar?. Dar estos valores de una forma coherente con el sentido común.
- Si no podemos suponer las hipótesis de independencia condicional, ¿qué probabilidades deberíamos pedir al experto? ¿Cuántos valores son, en total?
- ¿Cómo podemos calcular la probabilidad conjunta a partir de las condicionadas?. Aplicando el teorema de factorización, indica cómo se calcularía la probabilidad de que el paciente fume dado que tiene disnea

Redes bayesianas

Índice

Modelado con redes bayesianas

El matrimonio es la principal causa de divorcio (Groucho Marx)

El proceso de modelado consta de:

- Modelado **cualitativo**
  - Nodos o Variables (deben tomar un conjunto **EXHAUSTIVO y EXCLUYENTE** de valores).
  - Relaciones (deben ser de **influencia causal**)
- Modelado **cuantitativo**
  - Parámetros:
    - Probabilidad condicionada de cada nodo dados sus padres
    - Probabilidad a priori de los nodos sin padre

Modelado con redes bayesianas

Ejemplo 1: un problema de diagnóstico

Considera la siguiente situación: Los padres de Luisito, que acaba de cumplir un año, deciden llevarlo al pediatra porque vomita con cierta frecuencia. Con el pediatra sostienen la siguiente conversación:

Pediatra -. Denme toda la información que consideren que puede ser relevante.

Mamá-. El otro día Luisito estaba resfriado. Vomitó el biberón de la noche, creo que por culpa de los mocos, ya que había muchos en el vómito. Otras veces parece que vomita por una pequeña indigestión.

Papá-. Además creo que debe saber que mi hermano es celíaco (Aclaración: la celiaquía es una intolerancia al gluten, que poco a poco hace que se destruya el vello intestinal. Los vómitos son uno de sus síntomas más relevantes. Se cree que tiene cierta componente hereditaria).

Pediatra-. ¿Y la dieta de Luisito incluye gluten?

Ambos-. Sí, desde hace unos meses.

Plantea este **problema de diagnóstico** mediante una red bayesiana

Modelado con redes bayesianas

Ejemplo 2: Luis va de visita

Una tarde, Luis va a visitar a su compañero de oficina Antonio, y de repente comienza a estornudar. Luis piensa que se ha resfriado. Pero de repente observa que los muebles de Antonio están arañados, de forma que se le ocurre que quizás su amigo tenga un gato y sus estornudos se deban a una crisis producida por una rinitis alérgica.

Modelado con redes bayesianas

Ejemplo 3: un problema de clasificación

En el planeta Zyx se pueden encontrar varias clases de animales, llamemos a estas clases Wurros, Hobexas y Wackas. Todos tienen un tamaño muy pequeño, y sus pieles son o bien escamosas o bien están cubiertas de suave pelo. Además, una observación atenta ha permitido deducir lo siguiente:

- Todos los Wurros tienen 5 ó 6 patas. Su color es rojizo, y tienen la piel peluda y suave.
- El número de patas de las Hobexas es un entero que varía uniformemente entre 4 y 6, ambos inclusive. Su piel es escamosa.
- En cuanto a las Wackas, tienen 4 ó 5 patas, y ofrecen a la vista una tonalidad casi siempre azul, pero a veces (20% de los casos) rojiza.
- Los animales que tienen un número impar de patas cojean siempre. Los animales que tienen un número par de patas cojean sólo cuando tienen alguna anomalía (malformación congénita, heridas, etc.), lo cual ocurre en el 10% de los casos para los animales de 4 patas, y en el 20% para los de seis.

Plantea el **problema de la clasificación** de animales de Zyx mediante una red bayesiana

Más sobre modelado:

Identificación de las variables:

Por ejemplo,en problemas de diagnóstico

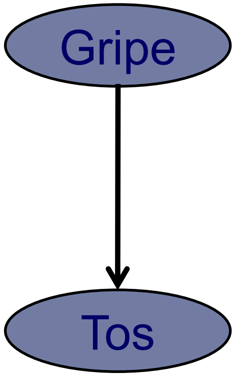
- ¿Cuál es la situación/problema que se plantea?
- ¿Qué posibles causas pueden explicar esta situación?
- ¿Qué otros factores pueden hacer que los problemas o causas ocurran, o impedir que ocurran?
- ¿De qué evidencia se dispone para soportar dichas causas, problemas o factores?

Más sobre modelado

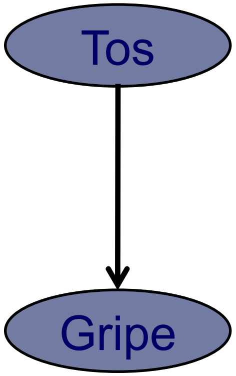
Tipo variable	Breve descripción
Objetivo	Modelan objetos de interés. No observables directamente.
Observación	Modelan la forma de medir variables objetivo. Pueden ser observadas directamente
Factor	Modelan fenómenos que afectan a otras variables del modelo
Promotor Inhibidor Requerido Preventivo	La variable afectada es más probable cuando están presentes. La variable afectada es menos probable cuando están presentes. Si no entra en acción, no ocurre la variable afectada. Si entra en acción, no ocurre la variable afectada.
Auxiliares	Usadas por conveniencia (para simplificar el modelo)

Más sobre modelado

Dos tipos de relaciones:



Relación causal



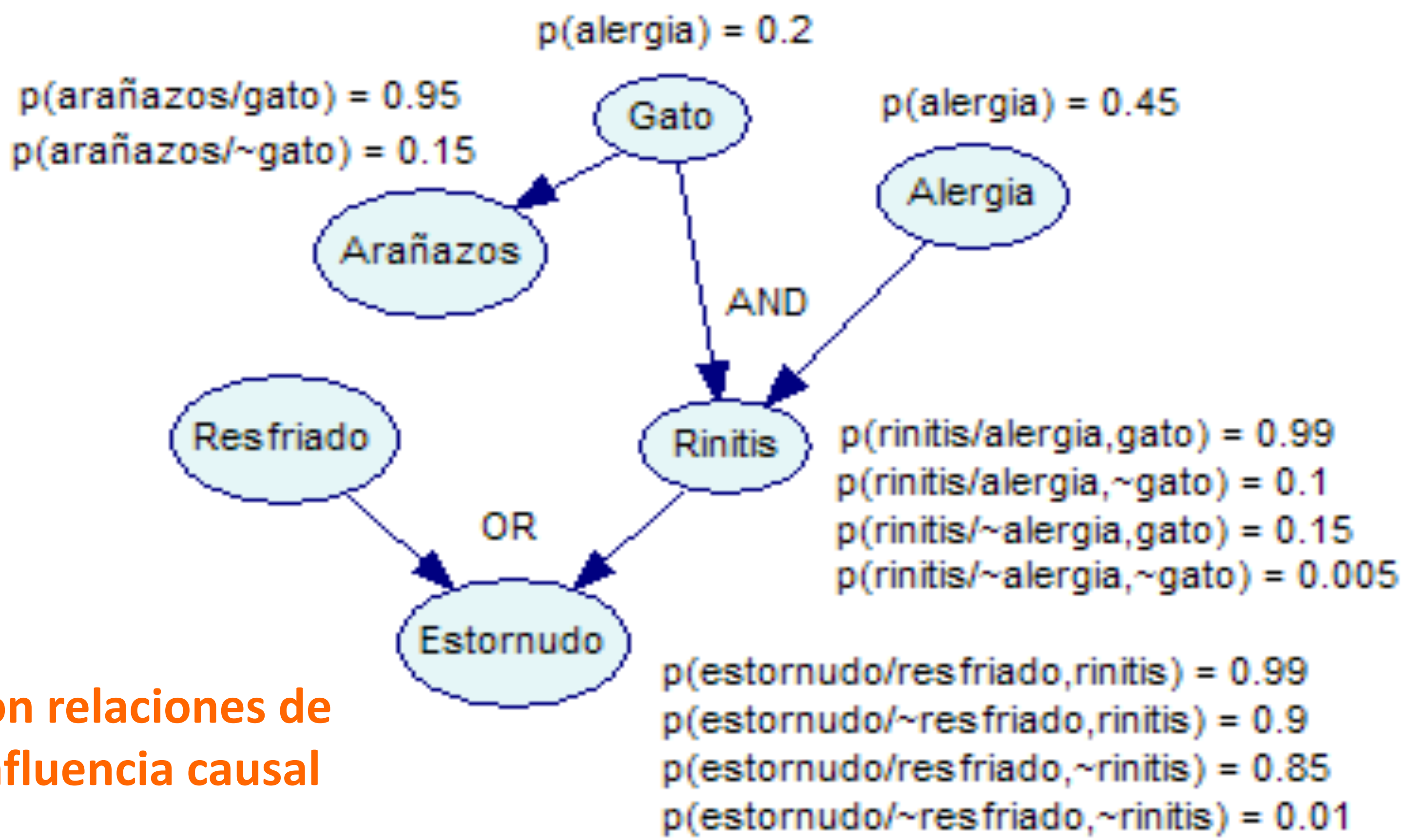
Regla de diagnóstico

Más sobre modelado

Si definimos los enlaces de forma que modelen la **relación causal**, el modelo obtenido es más sencillo de entender y menos complejo en cuanto a parámetros y relaciones.

Más sobre modelado

Ejemplo: el problema del estornudo



Con relaciones de influencia causal

Más sobre modelado

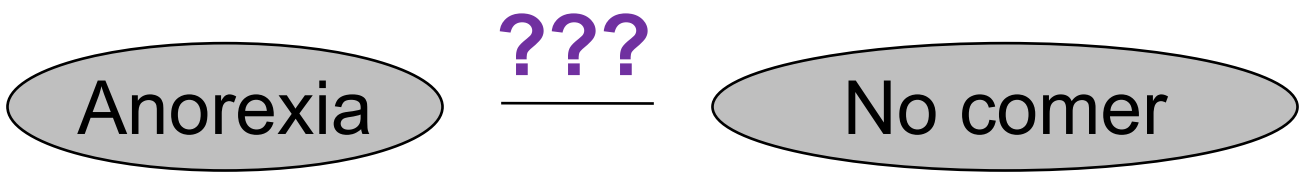
El problema del estornudo:

p(gato/arañazos,rinitis) = 0.920198
p(gato/arañazos,~rinitis) = 0.446214
p(gato/~arañazos,rinitis) = 0.0967387
p(gato/~arañazos,~rinitis) = 0.00742816

Con reglas de diagnóstico

Más sobre modelado

No siempre será fácil:



Más sobre modelado

Pero así son las cosas:

EL PAÍS.com | **Versión para imprimir** Imprimir

TRIBUNA: MANUEL LLORIS

Anorexia: causa o efecto

MANUEL LLORIS 26/07/2003

Leí días pasados en EL PAÍS el resumen de un informe de la profesora de la Complutense, Silvia Tubert. La anorexia, asegura esta experta, no es una enfermedad, sino un síntoma. Vino a verme un padre confundido y angustiado y tuve que sacarle de su error: no soy médico y me guardaré muy mucho de jugar a tal. Pensé luego que los expertos en ciertas cosas deberían ponerse de acuerdo con el fin de no sembrar alarmas en uno u otro sector de la sociedad. Personas anoréxicas, en general chicas cada vez más jóvenes, son causa de gran sufrimiento para sí mismas y para unos padres que ni siquiera saben a qué atenerse. La profesora Tubert remató la faena: "...el 90% de afectadas son mujeres porque detrás está el ideal de belleza que impone la sociedad". Reduccionismo que a mi visitante, padre de adolescente anoréxica, llenó de culpa (no había sabido educar) y a la vez de alivio, pues ahora toda la familia, de consuno, dirigiría a la víctima un incesante fuego graneado con el fin de arrasar de su conciencia este factor social. No soy médico -le insistí al padre- pero infórmese usted bien, no vaya a ser que agraven ustedes lo que pretenden sanar. Se despidió mohíno, más dolido y confuso que antes.

Como la anorexia produce delgadez extrema, pero no es enfermedad, sino síntoma, "engordar a los pacientes es secundario y supone taponar el problema". Uno creía que el síntoma (el efecto) puede en ocasiones constituirse en causa y ser causa más letal que la originaria. Puede que una persona, sometida a experimento científico, se quede en los huesos pero sin daño irreversible para su organismo. Vitaminas, minerales, aminoácidos y mínimo consumo energético. Admito que no conozco el dato, pero a falta de más información, no me parece inverosímil. Lo que sabemos todos es que la delgadez de la anoréxica no es un fenómeno controlado y que, por lo tanto, de prolongarse cierto tiempo puede dejar secuelas irreversibles de distinta gravedad. La desnutrición, ¿no es a la vez efecto y causa? Pero valga: no es enfermedad sino síntoma, así no lo entienda Zeus. Pero síntoma de qué. ¿De una disfunción metabólica? Eso, ¿no es una enfermedad? Todo fallo de adaptación al medio lo es, si bien podría aducirse, audazmente, que en casos de genialidad el fallo es el medio, incapaz de adaptarse al individuo genial

Más sobre modelado:

En cuanto a los parámetros:

- **Especificación directa** de los parámetros (con ayuda de expertos, costoso)
- **Aprendizaje** (si existen bases de datos)
  - *Aprendizaje de parámetros* (si se dispone de la estructura);
  - *Aprendizaje estructural* (se aprende tanto la estructura como los parámetros)
- **Combinaciones** de los dos anteriores (iterativamente)

Más sobre modelado (parámetros)

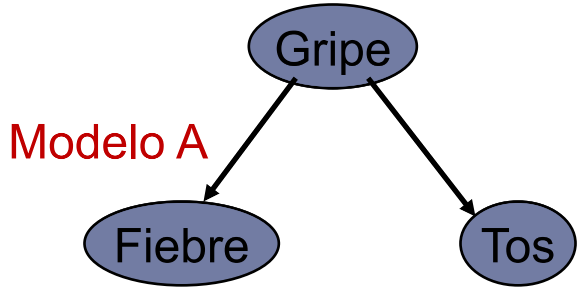
Si es necesario la especificación directa:

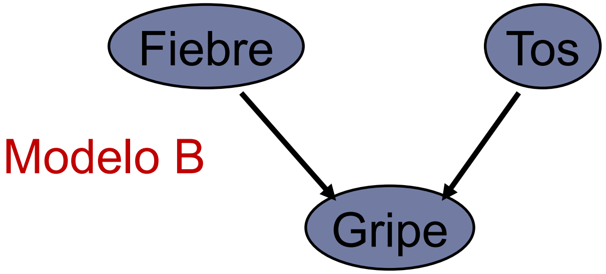
- En ausencia de información, se suele suponer equiprobabilidad
- En muchos casos, se podrán utilizar relaciones tipo AND u OR, con su correspondiente ruido

Algunos trucos útiles: verificación

Verificar las independencias del modelo:

(comprobar si las relaciones entre las variables reflejan adecuadamente las dependencias e independencias).



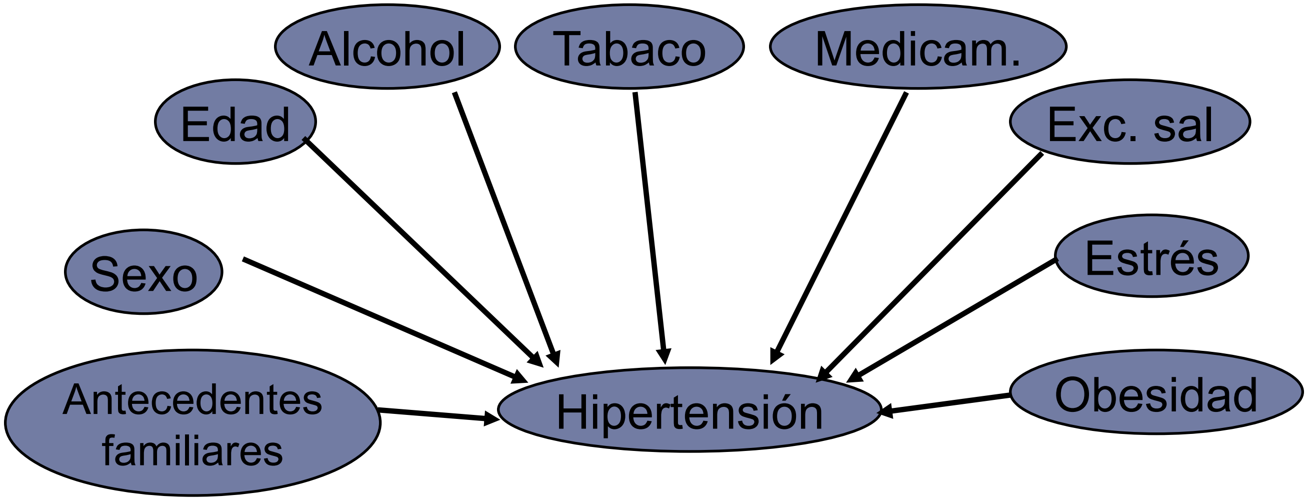


- En el modelo A, Fiebre y Tos son dependientes a priori pero independientes dado gripe
- En el modelo B, Fiebre y Tos son independientes a priori pero dependientes dado gripe (explaining-away)

Algunos trucos útiles: estructura

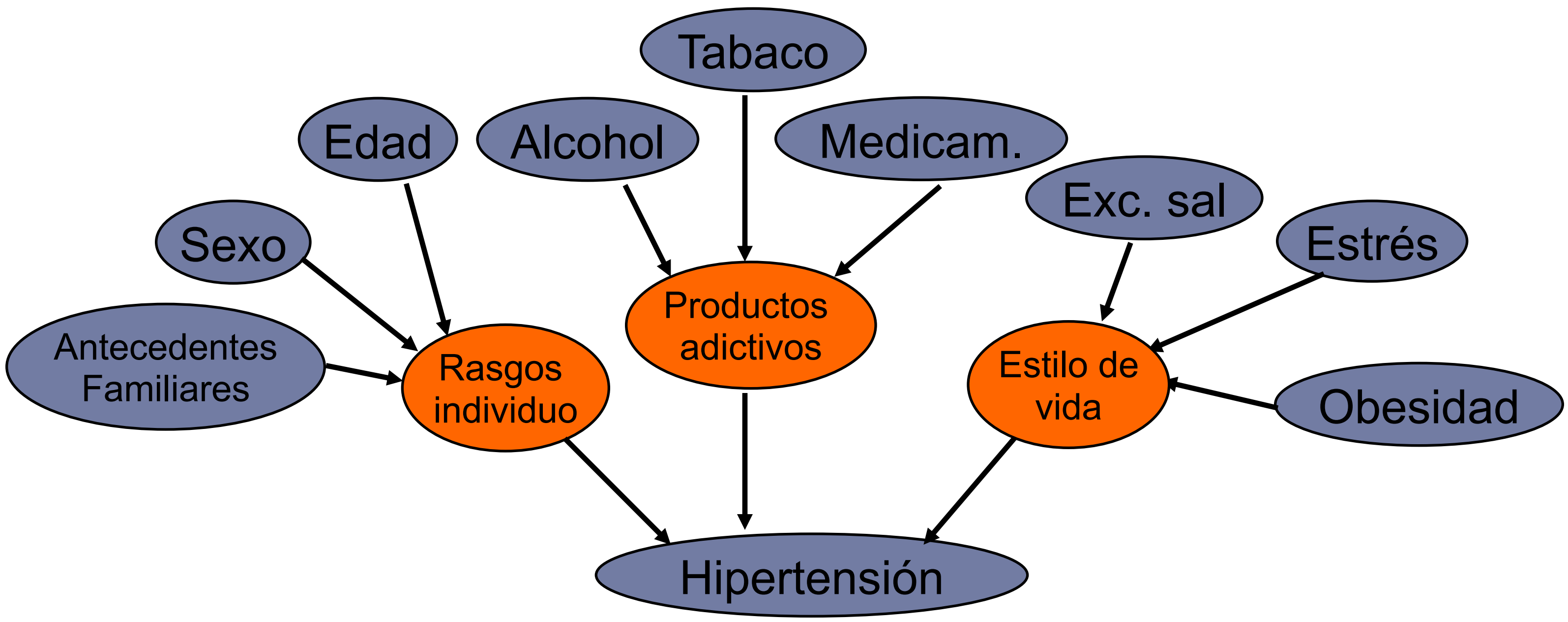
Introducir nodos intermedios

Es una buena idea para reducir la complejidad



Algunos trucos útiles: estructura

Si agrupamos nodos obtenemos un modelo más legible y menos complejo (en parámetros y en tiempo de ejecución de los algoritmos de propagación).



Algunos trucos útiles: parámetros

Utilizar modelos canónicos

- Son modelos que ayudan a simplificar la especificación de las tablas de probabilidades condicionadas, para ciertos tipos de relación causal entre variables.
- Los más utilizados son los modelos NOISY-OR y NOISY ADDER.

Algunos trucos útiles: parámetros

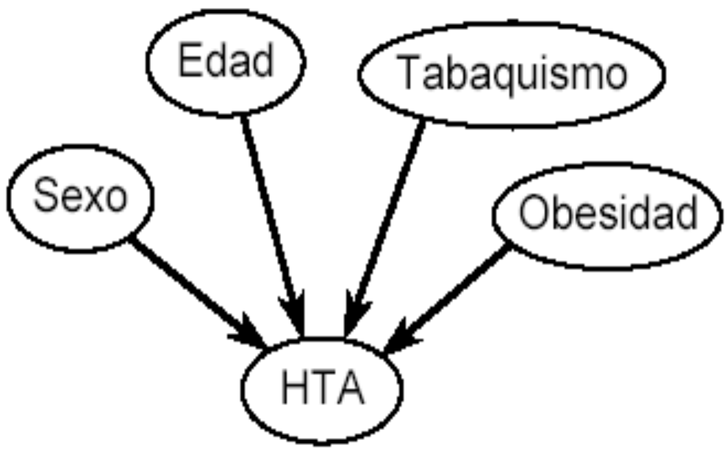
Modelo canónico: la puerta OR

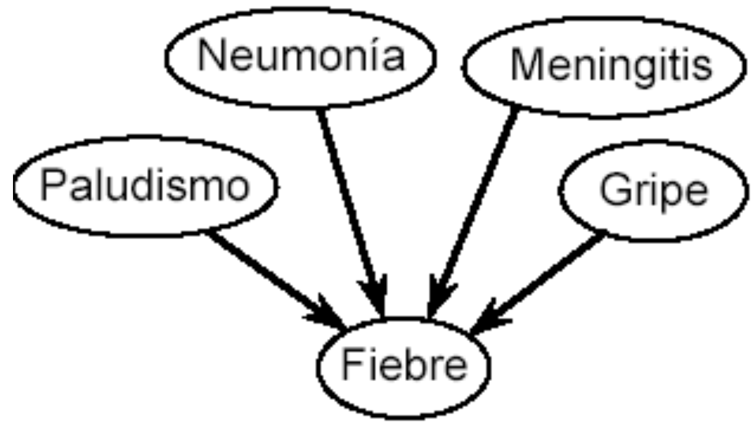
Hipótesis

1. Cada una de las causas, por sí misma, puede producir el efecto y basta con que una de ellas esté presente para que el efecto ocurra;
2. Cuando todas las causas están ausentes, el efecto está ausente;
3. No hay interacción entre las causas.

Algunos trucos útiles: parámetros

Ejemplos





Factores que tienen influencia causal en HTA:

No pueden modelarse con puerta OR

Factores que son causa de Fiebre:

Sí pueden modelarse con puerta OR

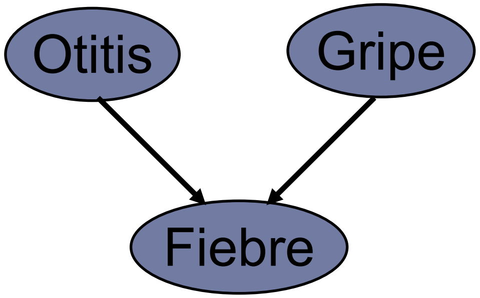
Algunos trucos útiles: parámetros

Si se cumplen las hipótesis de la puerta OR:

Sea c_x = Prob. de que causa x provoque el efecto

A partir de las c_x podemos calcular las probabilidades condicionadas

Veamos un ejemplo. Consideremos el modelo



Algunos trucos útiles: parámetros

Supongamos por ejemplo:

c_g = P(+f/+g)=0.8

c_o = P(+f/+o)=0.6

Entonces, las probabilidades condicionadas se pueden calcular así:

P(+f/+g, +o)= 0.8 + 0.2*0.6 = 0,92

P(+f/¬g, +o)= 0.6

P(+f/+g, ¬o)= 0.8

P(+f/ ¬g, ¬o)= 0

Algunos trucos útiles: parámetros

Modelo NOISY-OR (puerta OR ruidosa)

Supongamos que en el modelo queremos incluir que *es posible que otras causas no determinadas provoquen también fiebre*.

Entonces necesitamos los siguientes parámetros:

- Las probabilidades de que cada causa provoque el efecto por separado

  c_g = P(+f/+g, ¬o, ¬r)=0.8

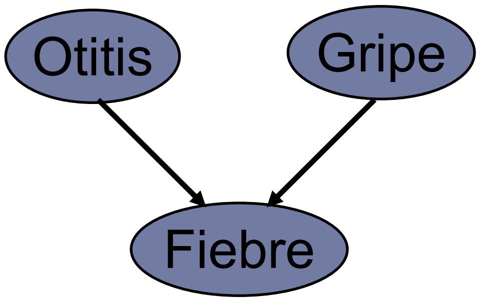
  c_o = P(+f/¬g, +o, ¬r)=0.6

- Un factor de ruido

  r = p(+f/¬g, ¬o)=0.01

Algunos trucos útiles: parámetros

Veamos como a partir de $c_{g,}$ c_o y r, se determinan las probabilidades necesarias para el modelo



- P(+f/+g,+o)
- P(+f/¬g,+o)
- P(+f/+g, ¬o)
- P(+f/¬g, ¬o) = r = 0,01

Algunos trucos útiles: parámetros

Por las hipótesis 1 y 2:

$P(+f/+g,+o,\neg r)= P(+f/+g,\neg o,\neg r) + P(\neg f/+g, \neg o, \neg r)P(+f/\neg g,+o,\neg r)= 0.8+0.2*0.6 =0,92$

Por la hipótesis 3:

$P(+f/\neg g, \neg o, \neg r) = 0$

Las otras probabilidades se calculan mediante:

$P(+f/+g,+o)= P(+f/+g,+o,\neg r)+P(\neg f/+g,+o,\neg r)P(+f/\neg g,\neg o)=0,92+0,08*0.01 = 0,9208$

$P(+f/+g,\neg o)= P(+f/+g,\neg o,\neg r) + P(\neg f/+g,\neg o,\neg r)P(+f/\neg g,\neg o) = 0,8 + 0.2*0.01 = 0,802$

$P(+f/\neg g, +o)= P(+f/\neg g,+o,\neg r) + P(\neg f/\neg g,+o,\neg r)P(+f/\neg g,\neg o) = 0,6 + 0.4*0.01 = 0,604$

Probabilidad de fiebre, dada cierta combinación de causas

+

Si dada esa combinación no hay fiebre, que actúe la causa no modelada (ruido)

Algunos trucos útiles: parámetros

La puerta OR para más de dos variables

En general, si en un modelo tenemos un efecto X y $U_1,\ldots,U_n$ son las causas posibles, si denotamos por c_i a la probabilidad que tiene cada causa de producir el efecto por separado y por q_i a la probabilidad complementaria ($q_i = 1-c_i$), entonces:

$$P(\neg x / u_1, \ldots., u_n) = \prod_{i \in T_u} q_i$$

donde Tu = conjunto de causas de X que están presentes.

Como en el caso anterior, el modelo se puede generalizar al caso de que haya cierto ruido (práctica 2)

Redes bayesianas

Índice

- Repaso de conceptos básicos de probabilidad
- Presentación intuitiva
- Definición formal de red bayesiana
- Teorema fundamental
- Modelado con redes bayesianas
- Algoritmo de propagación de probabilidades en árboles

Algoritmo de propagación de probabilidades

(para redes con forma de árbol)

Se desarrolla en dos fases:

Fase de inicialización

En esta fase se obtienen las probabilidades a priori de todos los nodos de la red, obteniendo un estado inicial de la red que denotaremos por S_0

Fase de actualización

Cuando una variable se instancia, se actualiza el estado de la red, obteniéndose las probabilidades a posteriori de las variables de la red basadas en la evidencia considerada, adoptando la red un estado que denotaremos por S_1

Este paso se repite cada vez que una variable se instancia, obteniéndose los sucesivos estados de la red

Algoritmo de propagación de probabilidades

(para redes con forma de árbol)

Idea principal:

Cada vez que una variable se instancia, informa a sus nodos vecinos mediante el paso de lo que llamaremos mensajes, de la siguiente forma:

- La variable envía a su padre un mensaje, que llamaremos el **λ-mensaje**, para informarle de que ha cambiado su valor. En base a este mensaje, se calcula un **λ-valor** para el nodo que lo recibe
- La variable envía a todos sus hijos un mensaje, que llamaremos el **π-mensaje**, para informarlos de que ha cambiado su valor. En base a este mensaje, se calcula un **π-valor** para el nodo que lo recibe

Las probabilidades a posteriori de cada nodo se obtienen multiplicando los λ-valores por los π-valores y normalizando

Así, la información se va propagando por la red tanto en **sentido ascendente como descendente**

Algoritmo de propagación: fórmulas

- **λ-mensaje** de B a A $\qquad \lambda_B(a_j) = \sum_{i=1}^{k} P(b_i/a_j)\lambda(b_i)$

- **π-mensaje** de A a B $\qquad \pi_B(a_j) = \begin{cases} \pi(a_j) \prod_{\substack{C \in s(A) \\ C \neq B}} \lambda_C(a_j) & \text{si A no ha sido instanciada (*)} \\ 1 & \text{si } A = a_j \\ 0 & \text{si } A \neq a_j \end{cases}$

- **λ-valor de *B*** $\qquad \lambda(b_i) = \begin{cases} \prod_{C \in s(B)} \lambda_C(b_i) & \text{si B no ha sido instanciada} \\ 1 & \text{si } B = b_i \\ 0 & \text{si } B \neq b_i \end{cases}$

- **π-valor de *B*** $\qquad \pi(b_i) = \sum_{j=1}^{m} P(b_i/a_j)\, \pi_B(a_j)$

- **Probabilidad de B:** $\qquad P^*(b_i) = \alpha\, \lambda(b_i)\, \pi(b_i)$

Algoritmo de propagación (II): Etapas

Etapa 1. Inicialización

A. Inicializar todos los λ-mensajes y λ-valores a 1

B. Si la raíz *A* tiene m posibles valores, entonces:

para $j = 1,\ldots,m, \quad \pi(a_j) = P(a_j)$

C. Para todos los hijos *B* de la raíz *A*, hacer

Enviar un nuevo π-mensaje a *B* usando la fórmula 2

(En el momento que una variable recibe un π-mensaje comenzará un flujo de propagación debido al procedimiento de actualización C)

Algoritmo de propagación: etapa de actualización

A. Si una variable *B* se instancia a un valor b_j, entonces

A.1. Inicializar $P^*(b_j) = 1$ y $P^*(b_i) = 0$, para todo $i \neq j$

A.2. Calcular $\lambda(B)$ usando la fórmula 3

A.3. Enviar un nuevo λ-mensaje al padre de *B* usando la fórmula 1

A.4. Enviar nuevos π–mensajes a los hijos de *B* usando la fórmula 2

B. Si una variable B recibe un nuevo λ-mensaje de uno de sus hijos y la variable B no ha sido instanciada todavía, entonces,

B.1. Calcular el nuevo valor de $\lambda(B)$ usando la fórmula 3

B.2. Calcular el nuevo valor de $P^*(B)$ usando la fórmula 5

B.3. Enviar un nuevo λ-mensaje al padre de *B* usando la fórmula 1

B.4. Enviar nuevos π–mensajes a los otros hijos de *B* usando fórm. 2

C. Si una variable *B* recibe un nuevo π-mensaje de su padre y la variable *B* no ha sido instanciada todavía, entonces:

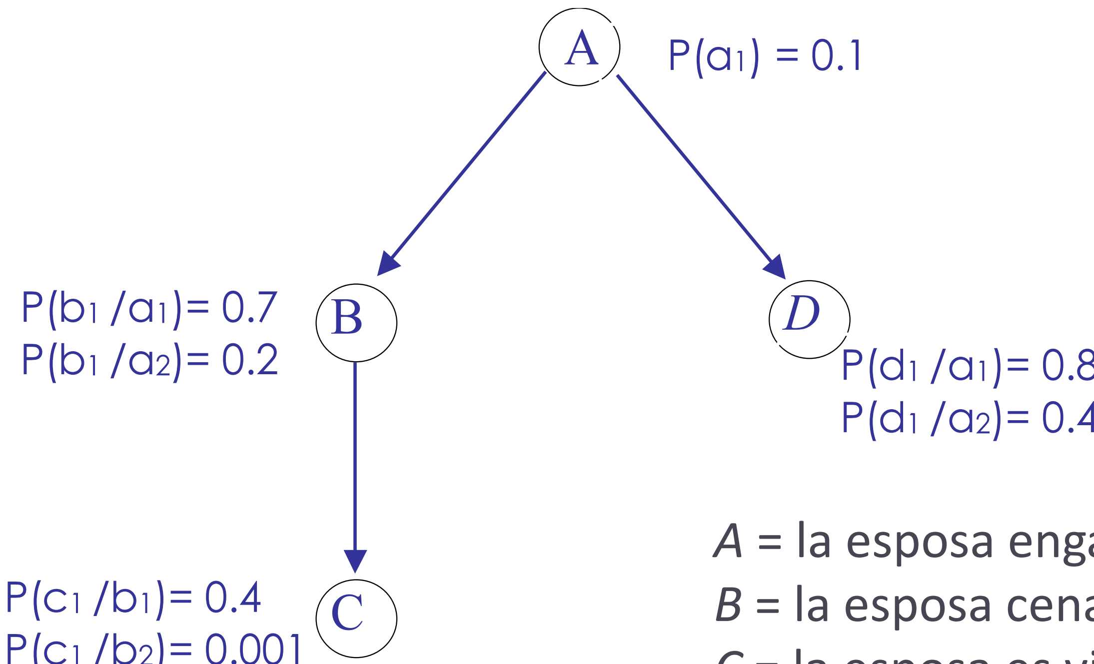
C.1. Calcular el nuevo valor de $\pi(B)$ usando la fórmula 4

C.2. Calcular el nuevo valor de $P^*(B)$ usando la fórmula 5

C.3. Enviar nuevos π–mensajes a los hijos de *B* usando la fórmula 2

Ejemplo

Supongamos que un señor piensa que su esposa le está siendo infiel. La red bayesiana que se construye para evaluar esta posibilidad es la siguiente:



A = la esposa engaña al marido

B = la esposa cena con otro

C = la esposa es vista cenando con otro

D = en el domicilio se reciben llamadas telefónicas sospechosas

Ejemplo: Cálculo probabilidades a priori

A. Ponemos todos los λ-mensajes y λ-valores a 1

B. Hacemos $\pi(a_j) = P(a_j)$, para $j = 1,2$. π(A)= (0.1, 0.9)

C. *A* envía un mensaje a su hijo, *B*,

$$\pi_B(a_1) = \pi(a_1)\lambda_D(a_1) = 0.1\ ;\ \pi_B(a_2) = \pi(a_2)\lambda_D(a_2) = 0.9$$

B toma entonces nuevos π-valores;

$$\pi(b_1) = P(b_1/a_1)\ \pi_B(a_1) + P(b_1/a_2)\ \pi_B(a_2) = 0.7\ 0.1 + 0.2\ 0.9 = 0.25$$

$$\pi(b_2) = P(b_2/a_1)\ \pi_B(a_1) + P(b_2/a_2)\ \pi_B(a_2) = 0.75$$

Y con ellos y con los λ-valores de B, se obtienen las probabilidades:

$$P(b_1) = \alpha\ 0.25\ 1 = 0.25\ ; \qquad P(b_2) = \alpha\ 0.75\ 1 = 0.75$$

Ahora, C recibe un π-mensaje por ser hijo de *B*:

$$\pi_C(b_1) = \pi(b_1) = 0.25\ ;\ \pi_C(b_2) = \pi(b_2) = 0.75$$

Y actualiza su π-valor:

$$\pi(c_1) = P(c_1/b_1)\ \pi_C(b_1) + P(c_1/b_2)\ \pi_C(b_2) = 0.4\ 0.25 + 0.001\ 0.75 = 0.10075$$

$$\pi(c_2) = P(c_2/b_1)\ \pi_C(b_1) + P(c_2/b_2)\ \pi_C(b_2) = 0.89925$$

A partir de ellos, calculamos las probabilidades de C, multiplicando por los λ-valores y normalizando:

$$P(c_1) = 0.10075; \qquad P(c_2) = 0.89925$$

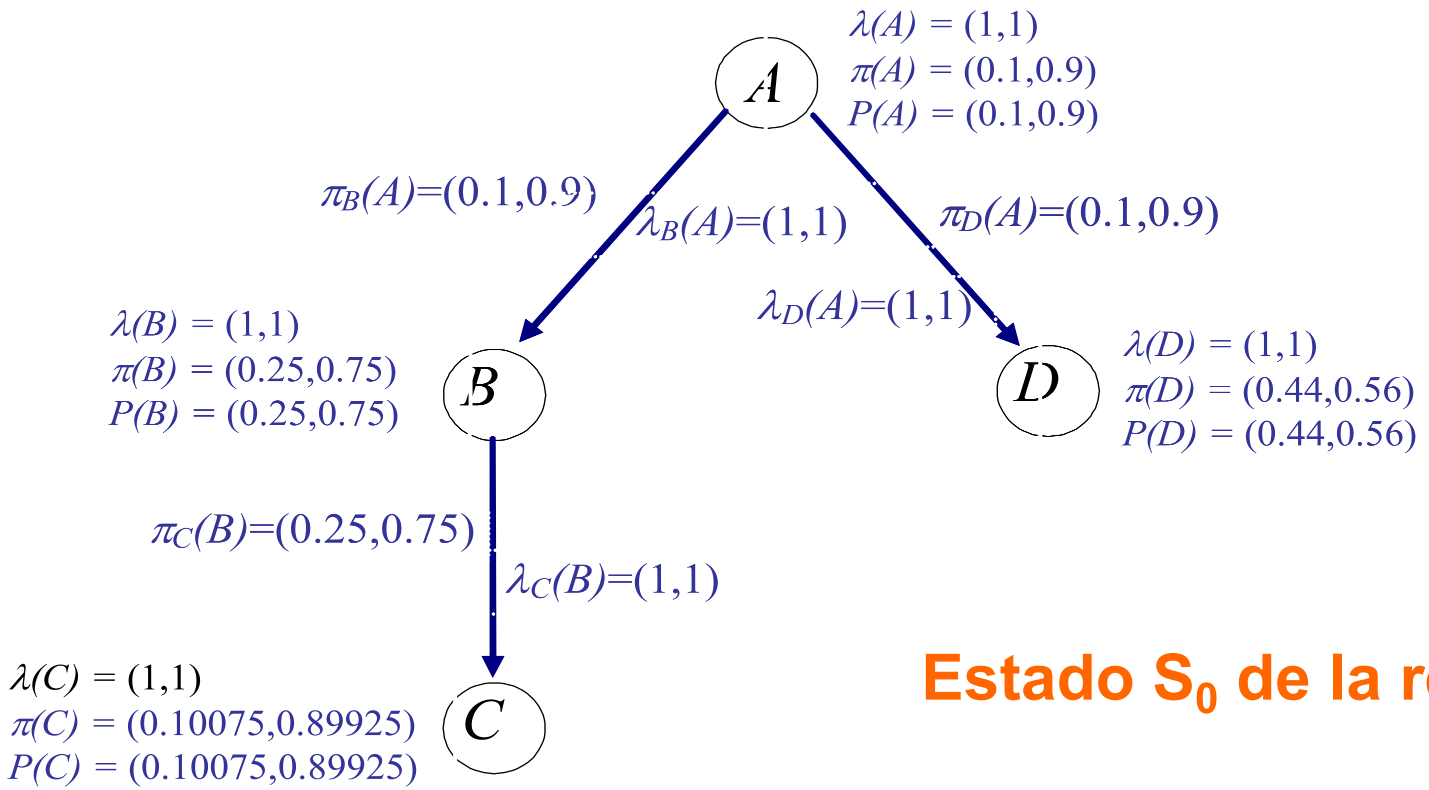
Students learn....

- 10% of what they read;
- 20% of what they hear;
- 30% of what they see;
- 50% of what they both see and hear;
- 70% of what they discuss with people whose opinions they value;
- 80% of what they personally experience; and
- 90% of what they teach to other people.

Treichler, D.G. (1967). Are you missing the boat in training aids? *Audio-visual communications*. New York: United Business Publications

Estado inicial S_0

Repitiendo el mismo procedimiento para D, obtenemos el estado inicial S_0:



Estado S_0 de la red

Propagación de evidencia (I)

Supongamos ahora que nos informan de que $B = b_1$

Actualización de B:

A.1 Calculamos ahora la probabilidad a posteriori de *B*, conocido que ha tomado el valor b_1, que evidentemente será:

$P^*(b_1) = 1$

$P^*(b_2) = 0$

A.2. Calculamos $\lambda(B)$;

$\lambda(b_1) = 1$

$\lambda(b_2) = 0$

A.3. Enviamos un λ-mensaje al padre de *B, A*

$\lambda_B(a_1) = P(b_1/a_1)\lambda(b_1) + P(b_2/a_1)\lambda(b_2) = 0.7\ 1 + 0.3\ 0 = 0.7$

$\lambda_B(a_2) = 0.2$

A.4. Enviamos un π-mensaje al hijo de *B, C*

$\pi_B(c_1) = 1$ puesto que *B* ha sido instanciada a b_1

$\pi_B(c_2) = 0$ puesto que *B* ha sido instanciada a b_1

Propagación de evidencia (I)

Actualización de C

Al recibir *C* un π-mensaje, se dispara el procedimiento de actualización *C*;

C.1. El π-valor de C cambia,

$$\pi(c_1) = P(c_1/b_1)\pi_C(b_1) + P(c_1/b_2)\,\pi_C(b_2) = 0.4$$

$$\pi(c_2) = 0.6$$

C.2. Calculamos la nueva probabilidad de *C*

$$P^*(c_1) = 0.4\,\alpha = 0.4$$

$$P^*(c_2) = 0.6\,\alpha = 0.6$$

C.3. No es necesario puesto que *C* no tiene hijos

Propagación de evidencia (II)

Actualización de A

Al recibir *A* un λ-mensaje, se dispara el procedimiento de actualización *B*;

B.1. Actualizamos el λ-valor

$$\lambda(a_1) = \lambda_B(a_1)\ \lambda_D(a_1) = 0.7$$
$$\lambda(a_2) = \lambda_B(a_2)\ \lambda_D(a_2) = 0.2$$

B.2. En base al λ-valor, calculamos la probabilidad a posteriori;

$$P^*(a_1) = \alpha\ 0.7\ \ 0.1 = 0.07\ \alpha = 0.28$$
$$P^*(a_2) = \alpha\ 0.2\ 0.9 = 0.18\ \alpha = 0.72$$

B.3. *A* no tiene padre

B.4. A envía un π-mensaje a su hijo, *D,*

$$\pi_D(a_1) = \pi(a_1)\ \lambda_B(a_1) = 0.1\ 0.7 = 0.07$$
$$\pi_D(a_2) = \pi(a_2)\ \lambda_B(a_2) = 0.9\ 0.2 = 0.18$$

Propagación de evidencia (II)

Actualización de D:

La variable *D* ha recibido un π-mensaje y por tanto debe actualizar su probabilidad. Aplicamos el procedimiento de actualización C:

C.1. El π-valor de *D* cambia,

$\pi(d_1) = 0.128$

$\pi(d_2) = 0.122$

C.2. Calculamos la nueva probabilidad de *D*

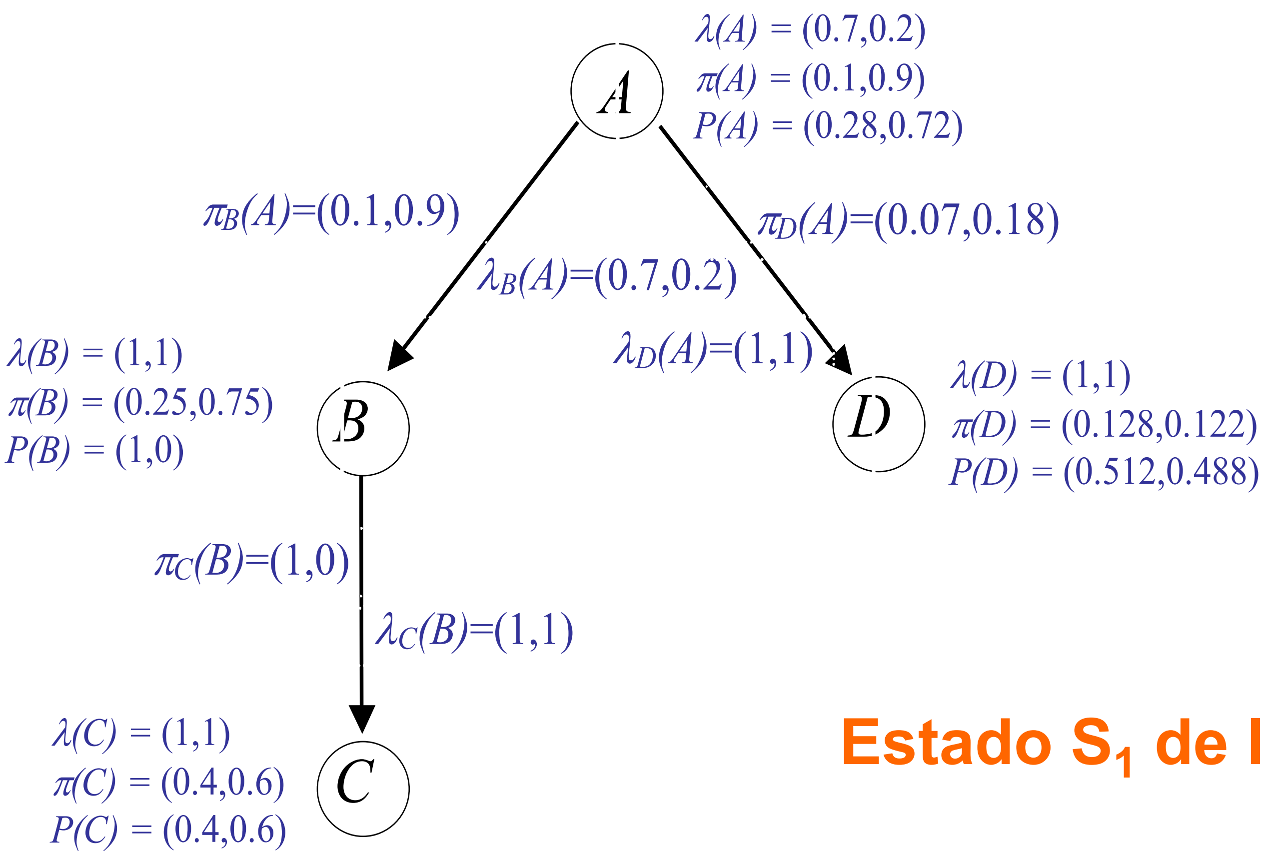
$P^*(d_1) = 0.512$

$P^*(d_2) = 0.488$

C.3. No es necesario puesto que *D* no tiene hijos

Estado S_1

Así, tras la instanciación de *B* a b_1, la red queda:



Estado S_1 de la red

Otra actualización

Supongamos ahora que la nueva evidencia es $e=\{D = d_2\}$

- *D* enviará un λ-mensaje a su padre, *A*,
- *A* enviará un π-mensaje a su hijo, *B*

B es un ***extremo muerto*** donde la propagación se para (en el caso de la propagación en árboles)

Terminar como problema propuesto

Evolución de las probabilidades

Estado inicial	Estado S_1	Estado S_2
$P(a_1)=0,1$	$P^*(a_1)=0,28$	$P^*(a_1)=0,1148$
$P(b_1)=0,25$	$P^*(b_1)=1$	$P^*(b_1)=1$
$P(c_1)=0,10075$	$P(^*c_1)=0,4$	$P^*(c_1)=0,4$
$P(d_1)=0,44$	$P^*(d_1)=0,512$	$P^*(d_1)=0$