Modelos Gráficos Probabilísticos

Tema 3. Construcción de redes bayesianas (continuación)

Andrés Cano Utrera

UNIVERSIDAD DE GRANADA

ETSIIT
Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

DECSAI

Curso 2024-2025

Referencias Básicas

Referencias Básicas

- E. Castillo, J.M. Gutiérrez y A.S. Hadi (1997). *Sistemas Expertos y Modelos de Redes Probabilísticos*. Academía de Ingeniería, Madrid. (disponible libremente en `http://personales.unican.es/gutierjm/papers/BookCGH.pdf`)
- D. Koller, N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, MA.
- F.V. Jensen, T.D. Nielsen (2007). *Bayesian Networks and Decision Graphs* (2nd. Edition). Springer-Verlag, Nueva York. (disponible en pdf desde red de la UGR `https://link.springer.com/book/10.1007/978-0-387-68282-2`)

Índice

Contenido del tema

Notación

- La descripción de un problema o situación dada se hará en términos de variables.
- Cada variable se notará con una letra mayúscula como $X, Y, \ldots$.
- Consideraremos variables categóricas, es decir que tienen un conjunto finito de valores posibles. El conjunto de valores posibles de Y se denota como U_Y. Un valor genérico de la variable Y se denotará como y (en minúscula).
- Un conjunto de variables o variable multidimensional se denotará habitualmente en negrita $\mathbf{X}, \mathbf{Y}, \ldots$. Se suele reservar $\mathbf{X}$ para el conjunto de todas las variables del problema estudiado. El conjunto de valores posibles de $\mathbf{Y}$ se denota como $U_{\mathbf{Y}}$.
- Un valor genérico de un conjunto $\mathbf{Y}$ se denota como $\mathbf{y}$.

Notación

- La descripción de un problema o situación dada se hará en términos de variables.
- Cada variable se notará con una letra mayúscula como $X, Y, \ldots$.
- Consideraremos variables categóricas, es decir que tienen un conjunto finito de valores posibles. El conjunto de valores posibles de Y se denota como U_Y. Un valor genérico de la variable Y se denotará como y (en minúscula).
- Un conjunto de variables o variable multidimensional se denotará habitualmente en negrita $\mathbf{X}, \mathbf{Y}, \ldots$. Se suele reservar $\mathbf{X}$ para el conjunto de todas las variables del problema estudiado. El conjunto de valores posibles de $\mathbf{Y}$ se denota como $U_{\mathbf{Y}}$.
- Un valor genérico de un conjunto $\mathbf{Y}$ se denota como $\mathbf{y}$.

Notación

- La descripción de un problema o situación dada se hará en términos de variables.
- Cada variable se notará con una letra mayúscula como $X, Y, \ldots$.
- Consideraremos variables categóricas, es decir que tienen un conjunto finito de valores posibles. El conjunto de valores posibles de Y se denota como U_Y. Un valor genérico de la variable Y se denotará como y (en minúscula).
- Un conjunto de variables o variable multidimensional se denotará habitualmente en negrita $\mathbf{X}, \mathbf{Y}, \ldots$. Se suele reservar $\mathbf{X}$ para el conjunto de todas las variables del problema estudiado. El conjunto de valores posibles de $\mathbf{Y}$ se denota como $U_{\mathbf{Y}}$.
- Un valor genérico de un conjunto $\mathbf{Y}$ se denota como $\mathbf{y}$.

Notación

- La descripción de un problema o situación dada se hará en términos de variables.
- Cada variable se notará con una letra mayúscula como $X, Y, \ldots$.
- Consideraremos variables categóricas, es decir que tienen un conjunto finito de valores posibles. El conjunto de valores posibles de Y se denota como U_Y. Un valor genérico de la variable Y se denotará como y (en minúscula).
- Un conjunto de variables o variable multidimensional se denotará habitualmente en negrita $\mathbf{X}, \mathbf{Y}, \ldots$. Se suele reservar $\mathbf{X}$ para el conjunto de todas las variables del problema estudiado. El conjunto de valores posibles de $\mathbf{Y}$ se denota como $U_{\mathbf{Y}}$.
- Un valor genérico de un conjunto $\mathbf{Y}$ se denota como $\mathbf{y}$.

Notación

- La descripción de un problema o situación dada se hará en términos de variables.
- Cada variable se notará con una letra mayúscula como $X, Y, \ldots$.
- Consideraremos variables categóricas, es decir que tienen un conjunto finito de valores posibles. El conjunto de valores posibles de Y se denota como U_Y. Un valor genérico de la variable Y se denotará como y (en minúscula).
- Un conjunto de variables o variable multidimensional se denotará habitualmente en negrita $\mathbf{X}, \mathbf{Y}, \ldots$. Se suele reservar $\mathbf{X}$ para el conjunto de todas las variables del problema estudiado. El conjunto de valores posibles de $\mathbf{Y}$ se denota como $U_{\mathbf{Y}}$.
- Un valor genérico de un conjunto $\mathbf{Y}$ se denota como $\mathbf{y}$.

El Teorema de Descomposición

Dada una red bayesiana con variables $\mathbf{X}$ entonces la distribución de probabilidad conjunta de estas variables se puede descomponer de la forma:

$$p(\mathbf{x}) = \prod_{y \in \mathbf{x}} p(y|pa(Y))$$

donde $pa(Y)$ es el conjunto de padres de la variable Y.

Consecuencia: Para especificar una red bayesiana solo hay que dar, para cada variable Y, una *distribución de probabilidad condicionada* (CPD) dados sus padres $pa(Y)$.

$$\sum_{y \in U_Y} p(y|pa(Y)) = 1$$

Si la variable es raíz, la distribución será la distribución marginal (sin condicionar ya que no tiene padres).

El Teorema de Descomposición

Dada una red bayesiana con variables $\mathbf{X}$ entonces la distribución de probabilidad conjunta de estas variables se puede descomponer de la forma:

$$p(\mathbf{x}) = \prod_{y \in \mathbf{x}} p(y|pa(Y))$$

donde $pa(Y)$ es el conjunto de padres de la variable Y.

Consecuencia: Para especificar una red bayesiana solo hay que dar, para cada variable Y, una *distribución de probabilidad condicionada* (CPD) dados sus padres $pa(Y)$.

$$\sum_{y \in U_Y} p(y|pa(Y)) = 1$$

Si la variable es raíz, la distribución será la distribución marginal (sin condicionar ya que no tiene padres).

El Teorema de Descomposición

Dada una red bayesiana con variables $\mathbf{X}$ entonces la distribución de probabilidad conjunta de estas variables se puede descomponer de la forma:

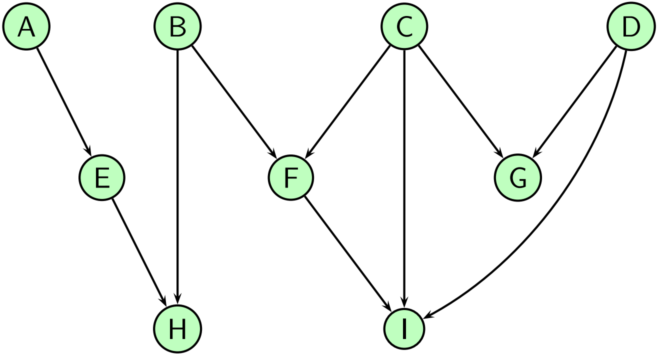
$$p(\mathbf{x}) = \prod_{y \in \mathbf{x}} p(y|pa(Y))$$

donde $pa(Y)$ es el conjunto de padres de la variable Y.
Consecuencia: Para especificar una red bayesiana solo hay que dar, para cada variable Y, una *distribución de probabilidad condicionada* (CPD) dados sus padres $pa(Y)$.

$$\sum_{y \in U_Y} p(y|pa(Y)) = 1$$

Si la variable es raíz, la distribución será la distribución marginal (sin condicionar ya que no tiene padres).

Ejemplo



$$p(a, b, c, d, e, f, g, h, i) =$$
$$p(a).p(b).p(c).p(d).p(e|a).p(f|b, c).p(g|c, d).p(h|b, e).p(i|c, d, f)$$

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \; \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1,\ldots,Y_k)$ especifica una distribución sobre X para cada asignación $y_1,\ldots,y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1,\ldots,Y_k)$ que verifique:

$$\phi(x, y_1,\ldots,y_k) = P(x|y_1,\ldots,y_k)\ \forall x, y_1,\ldots,y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \ \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \; \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \; \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \; \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \ \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \; \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Representación general de CPDs

- Una CPD $P(X|Y_1, \ldots, Y_k)$ especifica una distribución sobre X para cada asignación $y_1, \ldots, y_k$
- Para especificar una CPD, podemos usar cualquier función $\phi(X, Y_1, \ldots, Y_k)$ que verifique:

$$\phi(x, y_1, \ldots, y_k) = P(x|y_1, \ldots, y_k) \ \forall x, y_1, \ldots, y_k$$

- Ejemplos de posibles modelos:
  - Representación mediante tablas
  - CPDs deterministas
  - CPDs expresadas mediante árboles
  - CPDs expresados mediante reglas logísticas
  - Noisy OR/And
  - Modelo lineal gausiano

Elementos básicos: potenciales

Cada probabilidad condicional (por ejemplo $p(f|b,c)$) se puede ver como ejemplo de una entidad abstracta llamada potencial.

Potenciales

Un potencial consta de los siguientes elementos:

- Un conjunto de variables $\mathbf{Y}$.
- Una aplicación de $\mathbf{U_Y}$ en el conjunto de los reales R.

$p(f|b,c)$ es un potencial sobre $\{F,B,C\}$ y asigna a cada posible valor $F=f, B=b, C=c$ el valor $p(f|b,c)$.

Representación mediante tablas

- La forma más directa de representar un potencial es mediante una **tabla** (*array*) con tantas dimensiones como variables. Cuando una tabla se usa para representar una CPD, se suele usar el término Tabla de probabilidad condicional (CPT).
- Para cada dimensión correspondiente a una variable hay un índice con tantos valores como casos posibles tiene esta variable.
- Como veremos, aunque ésta sea la más obvia, no es la única forma de representar un potencial y, en determinadas circunstancias otras representaciones pueden ser más aconsejables.

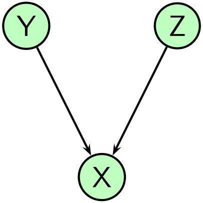
Representación mediante tablas

- La forma más directa de representar un potencial es mediante una **tabla** (*array*) con tantas dimensiones como variables. Cuando una tabla se usa para representar una CPD, se suele usar el término Tabla de probabilidad condicional (CPT).
- Para cada dimensión correspondiente a una variable hay un índice con tantos valores como casos posibles tiene esta variable.
- Como veremos, aunque ésta sea la más obvia, no es la única forma de representar un potencial y, en determinadas circunstancias otras representaciones pueden ser más aconsejables.

Representación mediante tablas

- La forma más directa de representar un potencial es mediante una **tabla** (*array*) con tantas dimensiones como variables. Cuando una tabla se usa para representar una CPD, se suele usar el término Tabla de probabilidad condicional (CPT).
- Para cada dimensión correspondiente a una variable hay un índice con tantos valores como casos posibles tiene esta variable.
- Como veremos, aunque ésta sea la más obvia, no es la única forma de representar un potencial y, en determinadas circunstancias otras representaciones pueden ser más aconsejables.

Representación mediante tablas: Ejemplo



Una distribución de probabilidad condicionada sobre X *Cáncer de Pulmón* condicionada a las variables Y *Fumador* y Z *Sexo* se puede representar como un potencial con lista de variables: X, Y, Z y tabla:

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.5	0.4	0.2	0.1
X=No	0.5	0.6	0.8	0.9

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$

    O sea, con k padres, es de orden $O(2^k)$

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$
    O sea, con k padres, es de orden $O(2^k)$
    - Ejemplo 1: Una variable con 5 variables padre binarias necesita de valores.
      Con 10 padres, necesitamos valores.
    - Ejemplo 2: Un síntoma (por ejemplo *Fiebre*) que dependa de 10 enfermedades.
      Sería muy tedioso preguntar al experto 1024 preguntas del tipo:
      ¿Cuál es la probabilidad de tener fiebre alta cuando un paciente tiene la enfermedad , y no tiene , ?

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$
    O sea, con k padres, es de orden $O(2^k)$
    - Ejemplo 1: Una variable X con 5 variables padre binarias necesita de $2^6 = 64$ valores.
      Con 10 padres, necesitamos $2^{11} = 2048$ valores
    - Ejemplo 2: Un síntoma (por ejemplo *Fiebre*) que dependa de 10 enfermedades.
      Sería muy tedioso preguntar al experto 1024 preguntas del tipo: ¿Cual es la probabilidad de tener fiebre alta cuando un paciente tiene la enfermedad A, y no tiene B, ...?

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$
    O sea, con k padres, es de orden $O(2^k)$
    - Ejemplo 1: Una variable X con 5 variables padre binarias necesita de $2^6 = 64$ valores.
      Con 10 padres, necesitamos $2^{11} = 2048$ valores
    - Ejemplo 2: Un síntoma (por ejemplo *Fiebre*) que dependa de 10 enfermedades.
      Sería muy tedioso preguntar al experto 1024 preguntas del tipo: ¿Cual es la probabilidad de tener fiebre alta cuando un paciente tiene la enfermedad A, y no tiene B, ...?

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$
    O sea, con k padres, es de orden $O(2^k)$
    - Ejemplo 1: Una variable X con 5 variables padre binarias necesita de $2^6 = 64$ valores.
      Con 10 padres, necesitamos $2^{11} = 2048$ valores
    - Ejemplo 2: Un síntoma (por ejemplo *Fiebre*) que dependa de 10 enfermedades.
      Sería muy tedioso preguntar al experto 1024 preguntas del tipo:
      ¿Cual es la probabilidad de tener fiebre alta cuando un paciente tiene la enfermedad A, y no tiene B, ...?

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$
    O sea, con k padres, es de orden $O(2^k)$
    - Ejemplo 1: Una variable X con 5 variables padre binarias necesita de $2^6 = 64$ valores.
      Con 10 padres, necesitamos $2^{11} = 2048$ valores
    - Ejemplo 2: Un síntoma (por ejemplo *Fiebre*) que dependa de 10 enfermedades.
      Sería muy tedioso preguntar al experto 1024 preguntas del tipo:
      ¿Cual es la probabilidad de tener fiebre alta cuando un paciente tiene la enfermedad A, y no tiene B, ...?

Representación mediante tablas

- Cualquier CPD discreta puede representarse mediante tablas.
- Como veremos, esta representación puede usarse de forma natural en algoritmos de inferencia.
- Desventajas:
  - En variables con dominio infinito (por ej. variables continuas) no pueden usarse.
  - El número de parámetros necesario para describir una CPD es igual a $|U_X| \cdot |U_{pa(X)}|$
    O sea, con k padres, es de orden $O(2^k)$
    - Ejemplo 1: Una variable X con 5 variables padre binarias necesita de $2^6 = 64$ valores.
      Con 10 padres, necesitamos $2^{11} = 2048$ valores
    - Ejemplo 2: Un síntoma (por ejemplo *Fiebre*) que dependa de 10 enfermedades.
      Sería muy tedioso preguntar al experto 1024 preguntas del tipo:
      ¿Cual es la probabilidad de tener fiebre alta cuando un paciente tiene la enfermedad A, y no tiene B, ...?

Representación de CPDs mediante una función determinista

Decimos que una CPD para X es una función determinista de sus padres $pa(X)$, cuando existe una función $f : U_{pa(X)} \rightarrow U_X$ tal que:

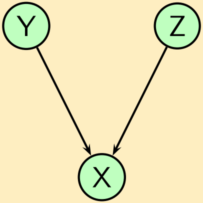
$$P(x|pa(X)) = \begin{cases} 1 & \text{si } \; x = f(pa(X)) \\ 0 & \text{en otro caso} \end{cases}$$

- O sea, todas probabilidades en la CPD serán 0.0 o 1.0.
- Cada una de ellas se puede obtener de forma indirecta a través de una función f.

Representación de CPDs mediante una función determinista

Ejemplo: Puerta OR

En un modelo que represente un circuito electrónico, la salida (X) de una puerta OR se puede modelar como una función OR determinista de las entradas (padres de X)



$$P(x|y,z) = \begin{cases} 1.0 & \text{si } x = y \text{ or } z \\ 0.0 & \text{en otro caso} \end{cases}$$

La representación como tabla sería:

	Y= 0	Y = 0	Y = 1	Y = 1
	Z= 0	Z= 1	Z = 0	Z = 1
X=0	1.0	0.0	0.0	0.0
X=1	0.0	1.0	1.0	1.0

Representación de CPDs mediante una función determinista

Otro ejemplo

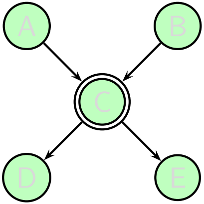
El genotipo de una persona se determina por dos copias de cada gen, llamados *alelos*. Cada alelo puede tomar uno entre varios valores. El fenotipo, a menudo, es una función determinista de estos valores. Por ejemplo, el gen responsable de determinar el **tipo de sangre** tiene tres valores: a, b y o. Sean G_1 y G_2 las variables que representan los dos alelos, y T la variable que representa el tipo de sangre en el fenotipo, entonces tenemos que:

$$T = \begin{cases} ab & \text{si } \ G_1 \text{ o } G_2 \text{ es } a \text{ y el otro es } b \\ a & \text{si al menos } \ G_1 \text{ o } G_2 \text{ es } a \text{ y el otro es } a \text{ o } o \\ b & \text{si al menos } \ G_1 \text{ o } G_2 \text{ es } b \text{ y el otro es } b \text{ o } o \\ 0 & \text{si } \ G_1 = o \text{ y } G_2 = o \end{cases}$$

Representación de CPDs mediante una función determinista

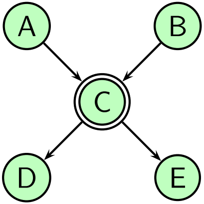
Se pueden obtener **más independencias** (de las proporcionadas por la D-separación) de la red bayesiana cuando tenemos nodos deterministas.

Ejemplo: Si conocemos el valor de A y de B, conoceremos el valor de C al ser éste un nodo determinista. Por tanto, D y E se hacen independientes al conocer A y B ($I(D, E|A, B)$).



Representación de CPDs mediante una función determinista

Se pueden obtener **más independencias** (de las proporcionadas por la D-separación) de la red bayesiana cuando tenemos nodos deterministas.
Ejemplo: Si conocemos el valor de A y de B, conoceremos el valor de C al ser éste un nodo determinista. Por tanto, D y E se hacen independientes al conocer A y B ($I(D, E|A, B)$).



Representación de CPDs mediante una función determinista

Algunas funciones deterministas pueden también implicar **otras independencias** aun más refinadas que ocurren para valores particulares de algunas variables.

Ejemplo: Consideremos en el ejemplo anterior que la función determinista para C es un OR. Supongamos que tenemos la evidencia $A = 1$. Puesto que C es un OR de sus padres, sabemos inmediatamente que $C = 1$ independientemente del valor de B. Por ello, podemos concluir que B y D son independientes si $A = 1$ y por ello:

$$P(D|B, A = 1) = P(D|A = 1)$$

Por otro lado, si $A = 0$, el valor de C no está determinado, y depende del valor de B.

Representación de CPDs mediante una función determinista

Algunas funciones deterministas pueden también implicar **otras independencias** aun más refinadas que ocurren para valores particulares de algunas variables.

Ejemplo: Consideremos en el ejemplo anterior que la función determinista para C es un OR. Supongamos que tenemos la evidencia $A = 1$. Puesto que C es un OR de sus padres, sabemos inmediatamente que $C = 1$ independientemente del valor de B. Por ello, podemos concluir que B y D son independientes si $A = 1$ y por ello:

$$P(D|B, A = 1) = P(D|A = 1)$$

Por otro lado, si $A = 0$, el valor de C no está determinado, y depende del valor de B.

Independencias basadas en el contexto

Independencia basada en el contexto

Sean $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ conjuntos disjuntos de variables, $\mathbf{C}$ un conjunto de variables (que puede solaparse con $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$), y sea $\mathbf{c} \in U_{\mathbf{C}}$. Decimos que $\mathbf{X}$ e $\mathbf{Y}$ son independientes contextualmente dado $\mathbf{Z}$ y el contexto $\mathbf{c}$ (denotado como $I_c(\mathbf{X}, \mathbf{Y} | \mathbf{Z}, \mathbf{c})$ si:

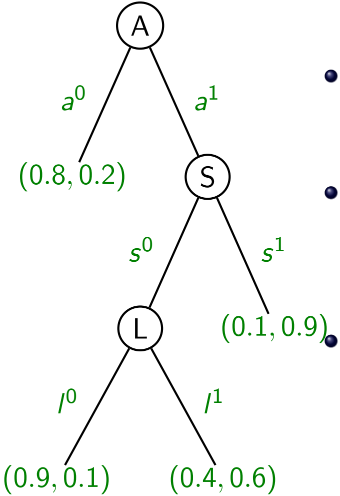
$$P(\mathbf{X}|\mathbf{Y}, \mathbf{Z}, \mathbf{c}) = P(\mathbf{X}|\mathbf{Z}, \mathbf{c}) \text{ siempre que } P(\mathbf{Y}, \mathbf{Z}, \mathbf{c}) > 0$$

o de forma equivalente

$$P(\mathbf{X}, \mathbf{Y}|\mathbf{Z}, \mathbf{c}) = P(\mathbf{X}|\mathbf{Z}, \mathbf{c}) \cdot P(\mathbf{Y}|\mathbf{Z}, \mathbf{c})$$

Representación de CPDs mediante árboles

Ejemplo: $P(X|A, S, L)$

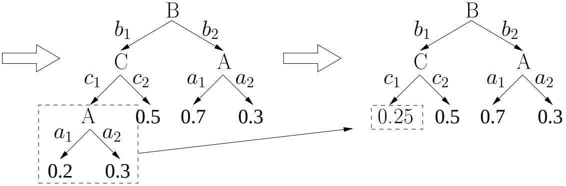


- A veces, una CPD para X contiene los mismos valores de probabilidad para varios valores de $pa(X)$.
- En general, a menudo encontramos que para una asignación parcial de valores $\mathbf{u}$ de las variables $\mathbf{U} \subset pa(X)$, los valores de los demás padres no son relevantes
- Los árboles permiten expresar independencias basadas en el contexto
  - $I_c(X, \{S, L\}|a^0)$
  - $I_c(X, \{L\}|\{a^1, s^1\})$

Representación de CPDs mediante árboles

- Algunos autores usan *árboles de probabilidad* para representar potenciales en general.
- En ese caso, es necesario incluir en el árbol la variable sobre la que se condiciona.
- Los árboles pueden ser podados para reducir su tamaño, aunque representarán el potencial de forma aproximada.

A	B	C	p(A,B,C)
a_1	b_1	c_1	0.2
a_1	b_1	c_2	0.5
a_1	b_2	c_1	0.7
a_1	b_2	c_2	0.7
a_2	b_1	c_1	0.3
a_2	b_1	c_2	0.5
a_2	b_2	c_1	0.3
a_2	b_2	c_2	0.3



Contenido del tema

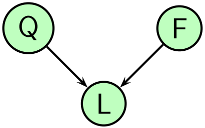
El modelo Noisy-Or

Ejemplo

Un profesor puede escribir una carta de recomendación para un estudiante (L). La calidad de la carta depende de dos factores:

- si el estudiante participó en clase, por ejemplo, haciendo buenas preguntas, (Q)
- y si escribió un buen trabajo final (F)

Sin embargo, el profesor puede olvidar la participación del estudiante en clase. También, podría no ser capaz de leer el trabajo (manuscrito) del estudiante, no apreciando su calidad. O sea, hay cierto ruido en el proceso.



El modelo Noisy-Or

Consideraremos cada una de las causas de forma aislada:

- Supongamos $P(l^1|q^1, f^0) = 0.8$: el profesor recuerda la participación del estudiante el 80% de las veces.
- $P(l^1|q^0, f^1) = 0.9$: el trabajo final del estudiante es legible en el 90% de los casos.
- Las dos causas son dos mecanismos causales independientes para provocar una buena carta de recomendación.
  - El primer mecanismo causal (participación en clase q^1) falla con probabilidad 0.2.
  - El segundo mecanismo (buen trabajo final f^1) falla con probabilidad 0.1.
  - En este modelo se asume que la probabilidad de que ambos mecanismos fallen se calcula como $0.2 \cdot 0.1$. Por tanto $P(l^0|q^1, f^1) = 0.02$ y $P(l^1|q^1, f^1) = 0.98$

El modelo Noisy-Or

De esta forma la CPD $P(L|Q, F)$ sería:

Q, F	l^0	l^1
$q^0 f^0$	1	0
$q^0 f^1$	0.1	0.9
$q^1 f^0$	0.2	0.8
$q^1 f^1$	0.02	0.98

Este tipo de interacción entre causas se conoce como modelo noisy-or

- Llamaremos λ_Q a la probabilidad de que Q cause L (0.8 en este caso): parámetro ruidoso
- Llamaremos λ_F a la probabilidad de que F cause L (0.9 en este caso): otro parámetro ruidoso
- Supongamos también una probabilidad residual $\lambda_0 = 0.0001$ que representa la probabilidad de que el profesor escriba una buena recomendación sin ninguna razón.

El modelo Noisy-Or

Podemos expresar las probabilidades de la anterior CPD usando los λ. Por ejemplo:

- Probabilidad de que no se de una buena carta de recomendación dado que el estudiante participó en clase e hizo un buen trabajo final:

$$P(l^0|q^1, f^1) = (1 - \lambda_0) \cdot (1 - \lambda_Q) \cdot (1 - \lambda_F) = 0.9999 \cdot 0.2 \cdot 0.1$$

- Probabilidad de que no se de una buena carta de recomendación dado que el estudiante no participó en clase e hizo un buen trabajo final:

$$P(l^0|q^0, f^1) = (1 - \lambda_0) \cdot (1 - \lambda_F) = 0.9999 \cdot 0.1$$

- Probabilidad de que no se de una buena carta de recomendación dado que el estudiante participó en clase y no hizo un buen trabajo final:

$$P(l^0|q^1, f^0) = (1 - \lambda_0) \cdot (1 - \lambda_Q) = 0.9999 \cdot 0.2$$

El modelo Noisy-Or

El modelo Noisy-Or

En general, si Y es una variable binaria con k padres binarios $X_1, \ldots, X_k$, la CPD $P(Y|X_1, \ldots, X_k)$ es un modelo noisy-or si hay $k+1$ parámetros ruidosos $\lambda_0, \lambda_1, \ldots, \lambda_k$ tal que:

$$P(y^0|X_1, \ldots X_k) = (1-\lambda_0) \prod_{i:X_i=x_i^1} (1-\lambda_i)$$

$$P(y^1|X_1, \ldots X_k) = 1 - [(1-\lambda_0) \prod_{i:X_i=x_i^1} (1-\lambda_i)]$$

El modelo Noisy-Or

El modelo Noisy-Or

Si interpretamos x_i^1 como 1 y x_i^0 como 0, podemos reescribir las ecuaciones anteriores de forma más compacta:

$$P(y^0|x_1,\ldots x_k) = (1-\lambda_0)\prod_{i=1}^{k}(1-\lambda_i)^{x_i}$$

Contenido del tema

Variables continuas

- En muchas situaciones, algunas variables se modelan mejor de forma que tomen valores en un dominio continuo: posición de un robot, velocidad, temperatura, presión, etc.
- Como se vio anteriormente, una solución podría ser la discretización de las variables continuas. Pero esto puede ser problemático:
  - Puede necesitarse de una discretización muy fina con decenas o cientos de valores para obtener un buen modelo.
    Por ejemplo, la posición de un robot debería discretizarse cada 15 centímetros para x e y. En un entorno de tamaño razonable, esto daría miles de valores para x y para y, dando lugar a millones de valores para la posición.
  - Al discretizar una variable continua, a menudo se pierde la estructura que la caracteriza. O sea, suposiciones de continuidad que ocurren en casi todos los dominios implican ciertas relaciones que ocurren entre las probabilidades asociadas con valores discretizados cercanos.

Variables continuas

- En muchas situaciones, algunas variables se modelan mejor de forma que tomen valores en un dominio continuo: posición de un robot, velocidad, temperatura, presión, etc.
- Como se vio anteriormente, una solución podría ser la discretización de las variables continuas. Pero esto puede ser problemático:
  - Puede necesitarse de una discretización muy fina con decenas o cientos de valores para obtener un buen modelo.
    Por ejemplo, la posición de un robot debería discretizarse cada 15 centímetros para x e y. En un entorno de tamaño razonable, esto daría miles de valores para x y para y, dando lugar a millones de valores para la posición.
  - Al discretizar una variable continua, a menudo se pierde la estructura que la caracteriza. O sea, suposiciones de continuidad que ocurren en casi todos los dominios implican ciertas relaciones que ocurren entre las probabilidades asociadas con valores discretizados cercanos.

Variables continuas

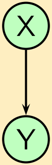
- En muchas situaciones, algunas variables se modelan mejor de forma que tomen valores en un dominio continuo: posición de un robot, velocidad, temperatura, presión, etc.
- Como se vio anteriormente, una solución podría ser la discretización de las variables continuas. Pero esto puede ser problemático:
  - Puede necesitarse de una discretización muy fina con decenas o cientos de valores para obtener un buen modelo.
    Por ejemplo, la posición de un robot debería discretizarse cada 15 centímetros para x e y. En un entorno de tamaño razonable, esto daría miles de valores para x y para y, dando lugar a millones de valores para la posición.
  - Al discretizar una variable continua, a menudo se pierde la estructura que la caracteriza. O sea, suposiciones de continuidad que ocurren en casi todos los dominios implican ciertas relaciones que ocurren entre las probabilidades asociadas con valores discretizados cercanos.

Variables continuas

- En muchas situaciones, algunas variables se modelan mejor de forma que tomen valores en un dominio continuo: posición de un robot, velocidad, temperatura, presión, etc.
- Como se vio anteriormente, una solución podría ser la discretización de las variables continuas. Pero esto puede ser problemático:
  - Puede necesitarse de una discretización muy fina con decenas o cientos de valores para obtener un buen modelo.
    Por ejemplo, la posición de un robot debería discretizarse cada 15 centímetros para x e y. En un entorno de tamaño razonable, esto daría miles de valores para x y para y, dando lugar a millones de valores para la posición.
  - Al discretizar una variable continua, a menudo se pierde la estructura que la caracteriza. O sea, suposiciones de continuidad que ocurren en casi todos los dominios implican ciertas relaciones que ocurren entre las probabilidades asociadas con valores discretizados cercanos.

Modelo lineal gausiano

Ejemplo del modelo lineal gausiano



Podemos modelar la CPD $P(Y|X)$ como una gausiana, cuyos parámetros dependen del valor de X.
Una solución muy común es que la media de Y sea una función lineal de X y que la varianza de Y no dependa de X. Por ejemplo:

$$p(Y|x) = \mathcal{N}(-2x + 0.9; 1)$$

Modelo lineal gausiano

Definición: Modelo lineal gausiano

Sea Y una variable continua con padres continuos $X_1, \ldots, X_k$. Decimos que Y sigue un modelo lineal gausiano si hay parámetros $\beta_0, \ldots, \beta_k$ y σ^2 tal que:

$$p(Y|x_1, \ldots x_k) = \mathcal{N}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k; \sigma^2)$$

O sea, Y es una función lineal de las variables padre, con la suma de un ruido gausiano (ϵ) de media 0 y varianza σ^2

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon$$

Red bayesiana gausiana

Son redes bayesianas en que todas las variables son continuas y definen sus CPDs a través del modelo lineal gausiano.

Modelo lineal gausiano

Definición: Modelo lineal gausiano

Sea Y una variable continua con padres continuos $X_1, \ldots, X_k$. Decimos que Y sigue un modelo lineal gausiano si hay parámetros $\beta_0, \ldots, \beta_k$ y σ^2 tal que:

$$p(Y|x_1, \ldots x_k) = \mathcal{N}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k; \sigma^2)$$

O sea, Y es una función lineal de las variables padre, con la suma de un ruido gausiano (ϵ) de media 0 y varianza σ^2

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon$$

Red bayesiana gausiana

Son redes bayesianas en que todas las variables son continuas y definen sus CPDs a través del modelo lineal gausiano.

Modelo condicional lineal gausiano (CLG)

Definición: Modelo condicional lineal gausiano (CLG)

Sea X una variable continua, y $\mathbf{U} = \{U_1, \ldots, U_m\}$ sus padres discretos e $\mathbf{Y} = \{Y_1, \ldots, Y_k\}$ sus padres continuos. Decimos que X sigue un modelo condicional lineal gausiano si para cada valor $\mathbf{u} \in Val(\mathbf{U})$ tenemos un conjunto de $k+1$ coeficientes $a_{\mathbf{u},0}, \ldots, a_{\mathbf{u},k}$ y una varianza $\sigma^2_{\mathbf{u}}$ tal que:

$$p(X|\mathbf{u}, \mathbf{y}) = \mathcal{N}(a_{\mathbf{u},0} + \sum_{i=1}^{k} a_{\mathbf{u},i} y_i; \sigma^2_{\mathbf{u}})$$

O sea, da una distribución lineal gausiana para cada configuración de los padres discretos $\mathbf{u}$

Definición: Red bayesiana condicional lineal gausiana

Red bayesiana en que todas las variables discretas tienen solo padres discretos y las variables continuas tienen asociada una CLG CPD.

Modelo condicional lineal gausiano (CLG)

Definición: Modelo condicional lineal gausiano (CLG)

Sea X una variable continua, y $\mathbf{U} = \{U_1, \ldots, U_m\}$ sus padres discretos e $\mathbf{Y} = \{Y_1, \ldots, Y_k\}$ sus padres continuos. Decimos que X sigue un modelo condicional lineal gausiano si para cada valor $\mathbf{u} \in Val(\mathbf{U})$ tenemos un conjunto de $k+1$ coeficientes $a_{\mathbf{u},0}, \ldots, a_{\mathbf{u},k}$ y una varianza $\sigma^2_{\mathbf{u}}$ tal que:

$$p(X|\mathbf{u}, \mathbf{y}) = \mathcal{N}(a_{\mathbf{u},0} + \sum_{i=1}^{k} a_{\mathbf{u},i} y_i; \sigma^2_{\mathbf{u}})$$

O sea, da una distribución lineal gausiana para cada configuración de los padres discretos $\mathbf{u}$

Definición: Red bayesiana condicional lineal gausiana

Red bayesiana en que todas las variables discretas tienen solo padres discretos y las variables continuas tienen asociada una CLG CPD.

Contenido del tema

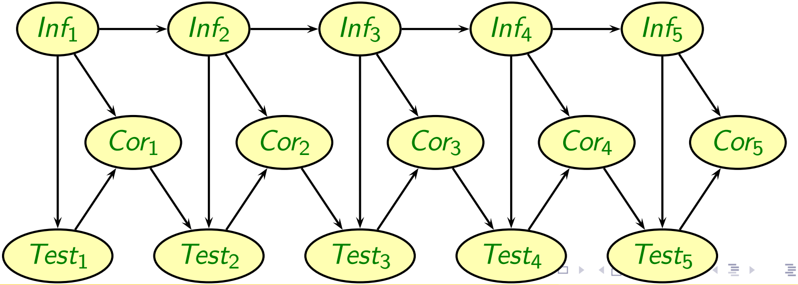
Modelos temporales

- Son útiles para dominios que evolucionan en el tiempo.
- Se define una descripción instantánea en un espacio de tiempo y se repite esta estructura en varias etapas de tiempo.
- Se considera que cada instante depende de los anteriores.

Modelos temporales

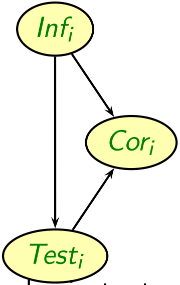
Por ejemplo, podemos tener una red que represente los análisis diarios (5 días) de la leche de una vaca (del libro de Jensen y Nielsen, pero simplificado).

- Inf_i: La leche está o no infectada (*si* o *no*).
- $Test_i$: Resultado del test para comprobar si la leche está infectada o no (*positivo* o *negativo*).
- Cor_i: El test i es correcto o no (*si* o *no*).



Modelos temporales

- En el anterior ejemplo, en cada instante i (cada día), tenemos tres variables: Inf_i, $Test_i$ y Cor_i
- Estas tres variables están conectadas en cada instante en la siguiente forma:



- Los enlaces entre variables de dos instantes se llaman enlaces temporales.

Modelos temporales

Modelo temporal

Contiene la misma estructura en cada instante y los enlaces temporales son siempre los mismos

Red dinámica bayesiana

Además tiene las mismas distribuciones de probabilidad condicional

Redes Dinámicas

La red bayesiana anterior puede definirse de forma compacta como una red dinámica con 5 instantes. Las flechas dobles indican relaciones temporales: