

Algoritmo de eliminación de variables - Factores

- Un factor corresponde a la probabilidad de un conjunto de variables dadas las variables ocultas
- Se representa por una tabla que para cada combinación de variables ocultas da la probabilidad de las variables del factor

$$f_X(Y, Z) =$$

Y	Z	
C	C	0.2
C	F	0.4
F	C	0.8
F	F	0.6

- Los factores tienen dos operaciones: suma y producto

Suma de Factores

- La suma se aplica a un factor y sobre una variable oculta del factor. Como resultado obtenemos una matriz reducida en la que las filas del mismo valor se han acumulado

$$f_{X\bar{Z}}(Y) = \sum_Z f_X(Y, Z) = \begin{array}{c|c} Y & \\ \hline C & 0.6 \\ F & 1.4 \end{array}$$

- Es igual que una operación de agregación sobre una columna en bases de datos

Producto de Factores

- El producto de factores permite juntar varios factores entre ellos utilizando las variables ocultas comunes

$$f_{X_1 X_2}(Y, W, Z) = f_{X_1}(Y, Z) \times f_{X_2}(Z, W) =$$

Y	Z		Z	W		Y	Z	W	
C	C	0.2	C	C	0.3	C	C	C	$0,2 \times 0,3$
C	F	0.8	C	F	0.7	C	C	F	$0,2 \times 0,7$
F	C	0.4	F	C	0.1	C	F	C	$0,8 \times 0,1$
F	F	0.6	F	F	0.9	C	F	F	$0,8 \times 0,9$
						F	C	C	$0,4 \times 0,3$
						F	C	F	$0,4 \times 0,7$
						F	F	C	$0,6 \times 0,9$
						F	F	F	$0,6 \times 0,3$

- Es igual que una operación de join en una base de datos multiplicando los valores de las columnas de datos

Algoritmo de eliminación de variables - ejemplo

Volveremos a calcular $P(\text{Fumador} | \text{Infarto} = \text{si}, \text{Deporte} = \text{no})$ a partir de la distribución de probabilidad conjunta:

$$P(D, A, S, F, I) = P(I | S, F) P(F) P(S | D, A) P(D) P(A)$$

Debemos calcular $P(F | I = \text{si}, D = \text{no})$, por lo tanto tenemos

$$\begin{aligned} P(F | I = s, D = n) &= \alpha P(I = s, F, D = n) \\ &= \alpha \sum_{A \in \{e, \neg e\}} \sum_{S \in \{a, n\}} P(D = n, A, S, F, I = s) \end{aligned}$$

En esta ocasión no sacamos factores comunes para seguir el algoritmo

$$\alpha P(D = n) \sum_{A \in \{e, \neg e\}} P(A) \sum_{S \in \{a, n\}} P(S | D = n, A) P(F) P(I = s | S, F)$$

Algoritmo de eliminación de variables - ejemplo

El algoritmo empieza calculando el factor para la variable *Infarto* ($P(I = s|S, F)$), esta tiene fijo su valor a si, depende de las variables *Presión Sanguinea* y *Fumador*

$$f_I(S, F) =$$

S	F	
a	s	0.8
a	n	0.7
n	s	0.6
n	n	0.3

La variable fumador ($P(F)$) no depende de ninguna otra variable, al ser la variable que preguntamos el factor incluye todos los valores

$$f_F(F) =$$

F	
s	0.4
n	0.6

Algoritmo de eliminación de variables - ejemplo

La variable *Presión Sanguinea* ($P(S|D = n, A)$), depende de las variable *Deporte* que tiene fijo su valor a no y *Alimentación*. Esta es una variable oculta, por lo que se debe calcular para todos sus valores

$$f_S(S, A) =$$

S	A	
a	e	0.25
a	$\neg e$	0.7
n	e	0.75
n	$\neg e$	0.3

Al ser la variable *Presión Sanguinea* una variable oculta debemos acumular todos los factores que hemos calculado

$$f_S(S, A) \times f_F(F) \times f_I(S, F)$$

Algoritmo de eliminación de variables - ejemplo

$$f_{FI}(S, F) = f_F(F) \times f_I(S, F) =$$

S	F	
a	s	0.8×0.4
a	n	0.7×0.6
n	s	0.6×0.4
n	n	0.3×0.6

$$f_{FIS}(S, F, A) = f_{FI}(S, F) \times f_S(S, A) =$$

S	F	A	
a	s	e	$0.8 \times 0.4 \times 0.25$
a	s	$\neg e$	$0.8 \times 0.4 \times 0.7$
a	n	e	$0.7 \times 0.6 \times 0.25$
a	n	$\neg e$	$0.7 \times 0.6 \times 0.7$
n	s	e	$0.6 \times 0.4 \times 0.75$
n	s	$\neg e$	$0.6 \times 0.4 \times 0.3$
n	n	e	$0.3 \times 0.6 \times 0.75$
n	n	$\neg e$	$0.3 \times 0.6 \times 0.3$

Algoritmo de eliminación de variables - ejemplo

Y ahora sumamos sobre todos los valores de la variable S para obtener el factor correspondiente a la variable *Presión Sanguinea*

$$f_{FIS}(F, A) = \sum_{S \in \{a, n\}} f_{FIS}(S, F, A) =$$

F	A	
s	e	$0.8 \times 0.4 \times 0.25 + 0.6 \times 0.4 \times 0.75 = 0.26$
s	$\neg e$	$0.8 \times 0.4 \times 0.7 + 0.6 \times 0.4 \times 0.3 = 0.296$
n	e	$0.7 \times 0.6 \times 0.25 + 0.3 \times 0.6 \times 0.75 = 0.24$
n	$\neg e$	$0.7 \times 0.6 \times 0.7 + 0.3 \times 0.6 \times 0.3 = 0.348$

Algoritmo de eliminación de variables - ejemplo

El factor de la variable *Alimentación* ($P(A)$) no depende de ninguna variable, al ser una variable oculta generamos todas las posibilidades

$$f_A(A) =$$

F	
e	0.4
$\neg e$	0.6

Ahora debemos acumular todos los factores calculados

$$f_{AFIS}(A) = f_A(A) \times f_{FIS}(F, A) =$$

F	A	
s	e	$0.26 \times 0.4 = 0.104$
s	$\neg e$	$0.296 \times 0.6 = 0.177$
n	e	$0.24 \times 0.4 = 0.096$
n	$\neg e$	$0.348 \times 0.6 = 0.208$

Algoritmo de eliminación de variables - ejemplo

Y ahora sumamos sobre todos los valores de la variable A para obtener el factor correspondiente a la variable *Alimentación*

$$f_{\overline{AFIS}}(F) = \sum_{A \in \{e, \neg e\}} f_{AFIS}(A) = \begin{array}{c|c} F & \\ \hline S & 0.104 + 0.177 = 0.281 \\ n & 0.096 + 0.208 = 0.304 \end{array}$$

Y por último la variable *Deporte* ($P(D = n)$) tiene el valor fijado a *no* y dado que no depende de la variable *fumador* se puede obviar, ya que es un factor constante.

Ahora, si normalizamos a 1

$$P(F|I = s, D = n) = \begin{array}{c|c} F & \\ \hline S & 0.48 \\ n & 0.52 \end{array}$$

Complejidad de la inferencia exacta

- La complejidad del algoritmo de eliminación de variables depende del tamaño del mayor factor, que depende del orden en el que se evalúan las variables y la topología de la red
- El orden de evaluación que escogeremos será el topológico según el grafo
- La complejidad de la inferencia exacta es NP-hard en el caso general
- Si la red bayesiana cumple que para cada par de nodos hay un único camino no dirigido (**poliárbol**) entonces se puede calcular en tiempo lineal
- Para obtener resultados en el caso general se recurre a algoritmos aproximados basados en técnicas de muestreo

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence.
- Topology + CPTs = compact representation of joint distribution.
- Generally easy for domain experts to construct.