Modelos de Bayes

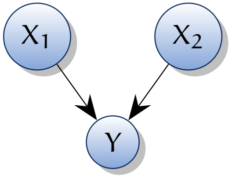
Verónica E. Arriola-Rios

Inteligencia Artificial

8 de junio de 2020

Temas

1. Independencia probabilística
2. Redes Bayesianas
   - Definición
   - Consultas
3. Inferencia Bayesiana
   - Flujo de influencia probabilística
   - Consultas



Eventos independientes

En el caso general, por la definición de probabilidad condicional:

$$P(\alpha \cap \beta) = P(\alpha|\beta)P(\beta) \tag{1}$$

Dos eventos α y β son independientes $P \models \alpha \perp \beta$[1] si:

$$P(\alpha \cap \beta) = P(\alpha)P(\beta) \tag{2}$$
$$P(\alpha|\beta) = P(\alpha) \tag{3}$$
$$P(\beta|\alpha) = P(\beta) \tag{4}$$

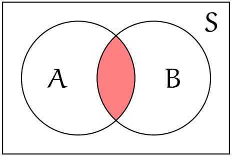


Figura: Intersección.

[1]Se lee: P satisface que α es independiente de β.

Variables independientes

Se dice que dos variables aleatorias A y B son *independientes* $P \models A \perp B$ si se cumple que:

$$P(A, B) = P(A)P(B) \tag{5}$$
$$P(A|B) = P(A) \tag{6}$$
$$P(B|A) = P(B) \tag{7}$$

Independencia $\neq$ exclusión

No hay forma de representar el concepto de independencia con diagramas de Venn. Obsérvese que independencia y exclusión ($A \cap B = \emptyset$) son dos conceptos distintos.

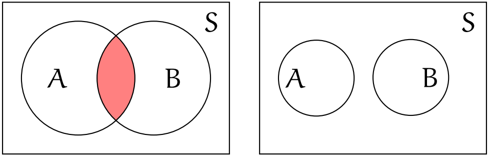


Figura: Intersección y exclusión.

Independencia condicional

- Dos eventos/variables aleatorias E_1 y E_2 son *condicionalmente independientes* dado F

$$P \models (E_1 \perp E_2|F) \tag{8}$$

- si, dado que F ocurre, la probabilidad condicional de que E_1 ocurra no cambia al obtenerse información sobre si E_2 ocurre o no.

$$\begin{aligned} P(E_1|E_2F) &= P(E_1|F) && (9) \\ P(E_2|E_1F) &= P(E_2|F) && (10) \\ & && (11) \end{aligned}$$

- o equivalentemente:

$$P(E_1E_2|F) = P(E_1|F)P(E_2|F) \tag{12}$$

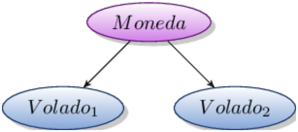
Ejemplo: Volados cargados

- Se tienen dos monedas.
  - Una cargada, el 90 % del tiempo cae sol.
  - Una normal.

  Esto define una primer variable aleatoria:

$$Moneda =< Cargada, Normal > \tag{13}$$

- Se realiza una serie de experimentos aleatorios:
  1. Se elige una moneda al azar.
  2. Se lanzan n volados $Volado_i =< Sol, \acute{A}guila >$ (aquí $n = 2$)
- Ya que el resultado de los volados depende de la moneda elegida, esto se ilustra:



Consultas

- ¿Cuál es la probabilidad de sacar alguna de las monedas?

$Moneda$	$P(Moneda)$
Cargada	
Normal	

$$\sum = 1$$

- ¿Cuál es la probabilidad de obtener sol o águila en un volado? Observemos que:
  - Esta probabilidad depende de la moneda que se haya elegido.
  - Cada volado es un evento diferente.
  - Lo que necesitamos primero es la distribución de probabilidad condicional. Esta es la misma para todos los volados.

$Moneda$	$Volado$	$P(Volado \mid Moneda)$
Cargada	Sol	
Cargada	Águila	
Normal	Sol	
Normal	Águila	

Marginales

- ¿Qué probabilidad tenemos de obtener sol o águila?
- Para obtener las probabilidades de los resultados de los volados, necesitamos tomar en cuenta la posibilidad de haber elegido cualquiera de las dos monedas.

$$P(\text{Volado}_1|\text{Moneda}) \qquad\qquad P(\text{Moneda})$$

- Se debe obtener la distribución:

Volado_1	$P(\text{Volado}_1)$
Sol	
Águila	

$$\sum = 1$$

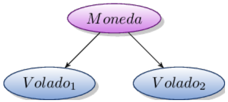
Posteriori

- La probabilidad de obtener Sol o $Águila$ en el segundo volado una vez que se conoce el resultado del primer volado $P(Volado_2|Volado_1)$. Ojo que la conexión se da porque fueron lanzados con la misma moneda.

$Volado_1$	$Volado_2$	$P(Volado_2\|Volado_1)$
Sol	Sol	
Sol	Águila	
Águila	Sol	
Águila	Águila	

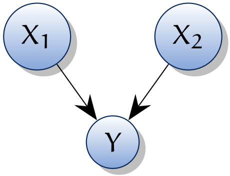
Análisis

- Inicialmente conocer el resultado del primer volado brinda información sobre la posibilidad de haber elegido la moneda cargada y por ende acerca del posible resultado del siguiente volado.
- Sin embargo, si se sabe qué moneda se eligió, esa información es suficiente para conocer la probabilidad de cada resultado en el segundo volado, la información del primer volado ya no es relevante.
- En este caso se dice que el resultado del segundo volado es independiente del resultado del primero, dado que se sabe qué moneda fue elegida.



Temas

1. Independencia probabilística
2. Redes Bayesianas
   - Definición
   - Consultas
3. Inferencia Bayesiana
   - Flujo de influencia probabilística
   - Consultas



Temas

1 Independencia probabilística

2 Redes Bayesianas
- Definición
- Consultas

3 Inferencia Bayesiana
- Flujo de influencia probabilística
- Consultas

Notación

De aquí en adelante seguiremos esta convención:

- Si la variable aleatoria se escribe con **mayúscula**, queremos los valores de la distribución de probabilidad *para todos sus valores posibles*.
- Si se escribe con **minúscula** nos referimos únicamente a *un valor en específico*.

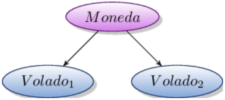
Red de Bayes

Definición (Red Bayesiana)

Una *Red Bayesiana* es:

- Una gráfica acíclica dirigida (GDA) G cuyos nodos representan a las variables aleatorias $X_1, ..., X_n$.
- Para cada nodo X_i define una distribución de probabilidad condicional

$$P(X_i | Padres_G(X_i)) \tag{14}$$

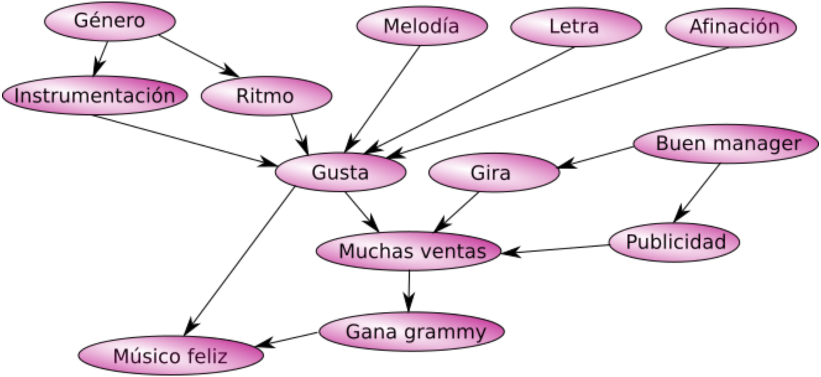


- La red de Bayes representa un **distribución de probabilidad conjunta** donde:

$$P(X_1, ..., X_n) = \prod_i P(X_i | Padres_G(X_i)) \qquad (15)$$

- A esta fórmula se le conoce como **regla de la cadena para Redes Bayesianas**.
- Cuando un distribución de probabilidad conjunta P satisface esta relación, se dice que P *se factoriza sobre* G.

Ejemplo: red del músico



Temas

1 Independencia probabilística

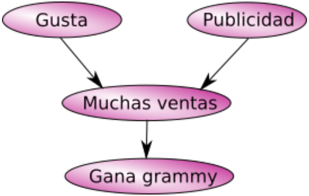
2 Redes Bayesianas
- Definición
- Consultas

3 Inferencia Bayesiana
- Flujo de influencia probabilística
- Consultas

Distribución de probabilidad conjunta completa

- Regla de la cadena:

$$
\begin{aligned}
&P(Gusta, Publicidad, Muchas\ ventas, Gana\ grammy) = \\
&\quad P(Gana\ grammy|Gusta, Publicidad, Muchas\ ventas) \\
&\quad P(Muchas\ ventas|Gusta, Publicidad) \\
&\quad P(Gusta|Publicidad) \\
&\quad P(Publicidad)
\end{aligned}
$$



Distribución de probabilidad conjunta completa

- Independencia condicional:

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

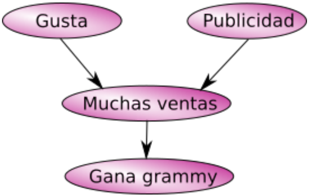
$$P(Gusta, Publicidad, Muchas\ ventas, Gana\ grammy) =$$

$$P(Gana\ grammy|\cancel{Gusta, Publicidad,} Muchas\ ventas)$$

$$P(Muchas\ ventas|Gusta, Publicidad)$$

$$P(Gusta|\cancel{Publicidad})$$

$$P(Publicidad)$$

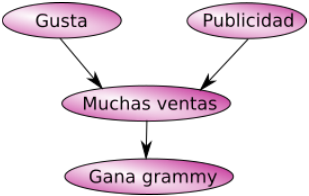


Distribución de probabilidad conjunta completa

- Regla de la cadena para redes de Bayes:

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

$$P(Gusta, Publicidad, Muchas\ ventas, Gana\ grammy) =$$
$$P(Gana\ grammy|Muchas\ ventas)$$
$$P(Muchas\ ventas|Gusta, Publicidad)$$
$$P(Gusta)$$
$$P(Publicidad)$$



Ejemplo

- Supongamos que tenemos tres variables con las distribuciones siguientes;

A	P(A)
a	0.7
aa	0.2
aaa	0.1

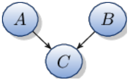
B	P(B)
b	0.33
bb	0.77

A	B	C	P(C|A,B)
a	b	0	0.2
a	b	1	0.8
aa	b	0	0.3
aa	b	1	0.7

...

- Para un solo renglón:

$$\begin{aligned} P(A = aa, B = b, C = 1) &= P(C = 1 | A = aa, B = b) P(A = aa) P(B = b) \\ &= P(1 | aa, b) P(aa) P(b) \\ &= 0.7 \times 0.2 \times 0.33 = 0.0462 \end{aligned}$$



Consultas: Reglas simples I

Teniendo la distribución de probabilidad conjunta completa:

- Se ha obtenido una tabla con:
  - Todas las variables del sistema.
  - Un renglón por cada combinación posible de los valores de las variables aleatorias.
- Para consultar cualquier maginal[2].
  - Marginalizar las variables que no se desean.

  Si sólo nos interesa un renglón con $\{x_1, ..., x_n\}$

$$P(x_1, ..., x_n) = \sum_{Y_1,\dots,Y_m} P(x_1, ..., x_n, Y_1, ..., Y_m)$$

  Si queremos la distribución completa:

$$P(X_1, ..., X_n) = \sum_{Y_1,\dots,Y_m} P(X_1, ..., X_n, Y_1, ..., Y_m)$$

[2] Distribución de probabilidad de un subconjunto de variables.

Marginal con factores

```
Factor:
['A', 'B', 'C', 'P(A,B,C)']
['a', 'b', 0, 0.0462]
['a', 'b', 1, 0.1848]
['a', 'bb', 0, 0.0]
['a', 'bb', 1, 0.5389999999999999]
['aa', 'b', 0, 0.0198]
['aa', 'b', 1, 0.0462]
['aa', 'bb', 0, 0.0]
['aa', 'bb', 1, 0.15400000000000003]
['aaa', 'b', 0, 0.0]
['aaa', 'b', 1, 0.033]
['aaa', 'bb', 0, 0.0]
['aaa', 'bb', 1, 0.07700000000000001]
```

```
P(B,C) Factor:
['B', 'C', 'P(B,C)']
['b', 0, 0.066]
['b', 1, 0.264]
['bb', 0, 0.0]
['bb', 1, 0.77]
```

```
P(B) Factor:
['B', 'P(B)']
['b', 0.33]
['bb', 0.77]
```

Consultas: Reglas simples II

Teniendo la distribución de probabilidad conjunta completa:

- Para obtener una distribución de probabilidad condicional:
  - Utilizar la definición de distribución de probabilidad condicional. Si sólo nos interesa un renglón con $\{x_1, ..., x_n\}$ para un caso de la evidencia $\{y_1, ..., y_m\}$:

$$P(x_1, ..., x_n | y_1, ..., y_m) = \frac{P(x_1, ..., x_n, y_1, ..., y_m)}{P(y_1, ..., y_m)}$$

  - Calcular el numerador y denominador según la regla anterior.
  - Realizar la división.

Distribución condicional con factores

- Dado el factor con $P(X_1, ..., X_n, Y_1, ..., Y_m)$, si queremos $P(X_1, ..., X_n | y_1, ..., y_m)$:
  1. Usar reducción de factores, quedándonos sólo con los renglones donde $\{Y_1 = y_1, ..., Y_m = y_m\}$
  2. Normalizar ese factor.
  3. Observar que la constante de normalización es, de hecho, $P(y_1, ..., y_m)$.

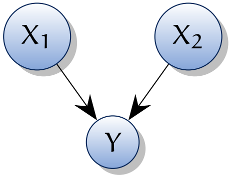
Condicional con factores

```
Factor:
['A', 'B', 'C', 'P(A,B,C)']
['a', 'b', 0, 0.0462]
['a', 'b', 1, 0.1848]
['a', 'bb', 0, 0.0]
['a', 'bb', 1, 0.538999999]
['aa', 'b', 0, 0.0198]
['aa', 'b', 1, 0.0462]
['aa', 'bb', 0, 0.0]
['aa', 'bb', 1, 0.15400000]
['aaa', 'b', 0, 0.0]
['aaa', 'b', 1, 0.033]
['aaa', 'bb', 0, 0.0]
['aaa', 'bb', 1, 0.0770000]
```

```
Factor:
['A', 'C', 'P(C,A,B='b')']
['a', 0, 0.0462]
['a', 1, 0.1848]
['aa', 0, 0.0198]
['aa', 1, 0.0462]
['aaa', 0, 0.0]
['aaa', 1, 0.033]
Factor:
['A', 'C', 'P(C,A|B='b')']
['a', 0, 0.14]
['a', 1, 0.56]
['aa', 0, 0.0600000000000]
['aa', 1, 0.14]
['aaa', 0, 0.0]
['aaa', 1, 0.100000000000]
```

Temas

1. Independencia probabilística
2. Redes Bayesianas
   - Definición
   - Consultas
3. Inferencia Bayesiana
   - Flujo de influencia probabilística
   - Consultas



Temas

1. Independencia probabilística
2. Redes Bayesianas
   - Definición
   - Consultas
3. Inferencia Bayesiana
   - Flujo de influencia probabilística
   - Consultas

Ejemplo de red de Bayes

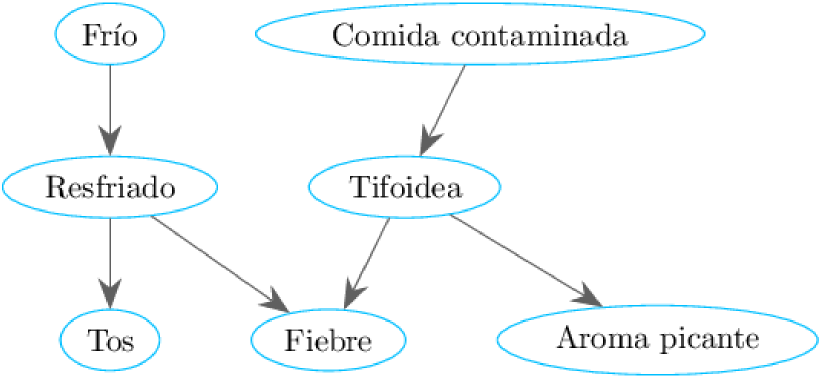
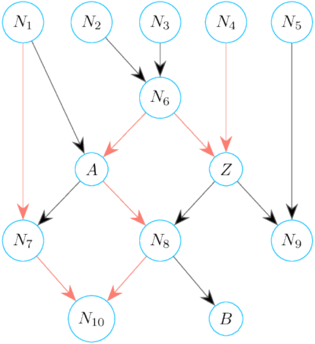


Figura: Construcción de una red de Bayes siguiendo relaciones de causalidad.

Definiciones

Definición (Ruta)

Una *ruta* $X_1 - \ldots - X_k$ es una secuencia de nodos que se encuentran conectados entre sí mediante una sola arista (no dirigida) en la gráfica.



Flujo de Información

Las reglas siguientes indican cuándo fluye información probabilística entre dos nodos A y B separados por un tercero Z, es decir cuándo esos nodos son probabilísticamente dependientes.

Se utiliza la clave siguiente:

Z	El valor de la variable Z no ha sido observado.
$\underline{Z}$	El valor de la variable Z ha sido observado.
$\longleftrightarrow$	Hay flujo de información entre las variables.
$\leftrightarrow$ x $\leftrightarrow$	El flujo se corta.

Ascendientes y descendientes

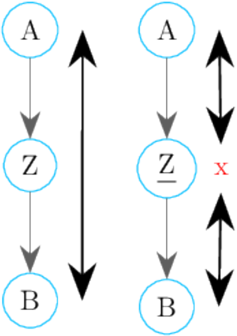


Figura: La información fluye entre ascendientes y descendientes siempre que no haya sido observado algún nodo en medio del camino.

Ancestro común

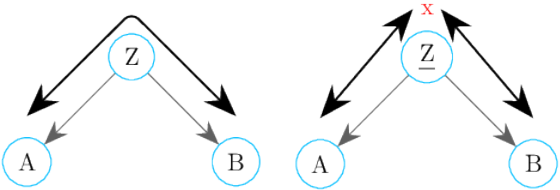


Figura: La información también fluye entre nodos hermanos, mientras el padre común no haya sido observado.

Estructura-V

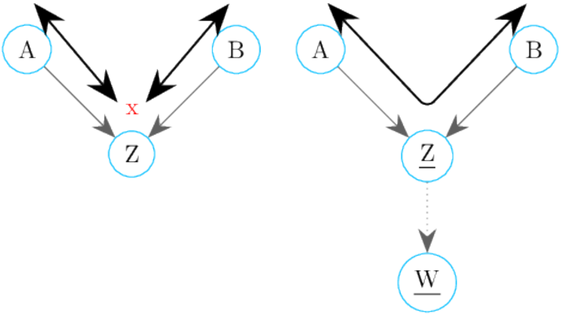


Figura: En el último caso, donde dos nodos son padres de un mismo nodo, el flujo se activa cuando el hijo común o algún descendiente suyo es observado; es costumbre referirse a la figura que forman los nodos $A - \underline{Z} - B$ como *estructura-v*.

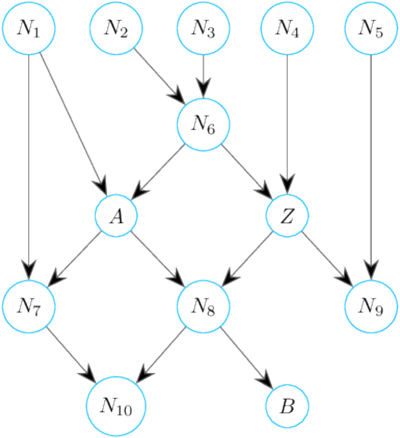
Ruta activa

Sea $\mathbb{Z}$ el conjunto de variables evidencia cuyo valor ha sido observado.

Definición

- Una ruta se encuentra *activa* si no tiene *estructuras-v*.
- Una ruta $X_1 - ... - X_k$ está *activa dado* $\mathbb{Z}$ si:
  - Para cualquier *estructura-v* $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ se tiene que X_i o alguno de sus descencientes están en $\mathbb{Z}$.
  - Ningún otro X_i está en $\mathbb{Z}$.

Ejercicio: rutas activas



D-Separación

Definición

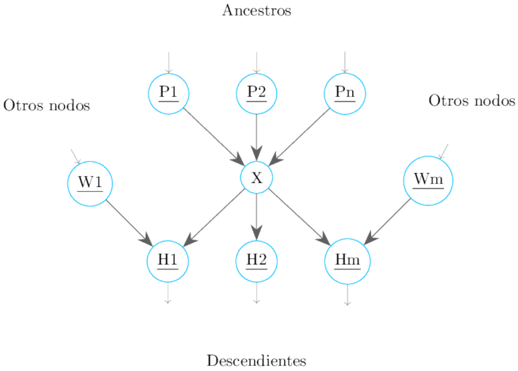
Dos variables aleatorias X y Y se encuentran *d-separadas* en G dado $\mathbb{Z}$ si no existe una ruta activa entre X y Y dado $\mathbb{Z}$.

$$\mathrm{d} - sep_{\mathrm{G}}(\mathrm{X}, \mathrm{Y}|\mathrm{Z}) \tag{16}$$

Manto de Márkov

Para que una variable quede aislada del resto de la red, se deben observar:

- Sus padres.
- Sus hijos.
- Los padres de sus hijos. Éstos porque, al observar a los hijos, se activó el flujo hacia ellos.



Temas

1 Independencia probabilística

2 Redes Bayesianas
- Definición
- Consultas

3 Inferencia Bayesiana
- Flujo de influencia probabilística
- Consultas

Consultas

Una consulta a una red bayesiana tendrá la forma del cálculo de una probabilidad.

- ¿Cuál es la probabilidad de que un paciente tenga tifoidea dado que presenta fiebre y despide un aroma picante? $P(Tifoidea = 1|Fiebre = 1, Aroma = 1)$
- ¿Cuál es la probabilidad de que un paciente tenga fiebre? $P(Fiebre = 1)$ ó $P(fiebre)$
- ¿Cuál es la probabilidad de que un paciente no tenga fiebre? $P(Fiebre = 0)$ ó $P(\neg fiebre)$
- ¿Cuál es la probabilidad de que un paciente tenga/no tenga fiebre? $P(Fiebre)$

- ¿Cuál es la probabilidad de que un paciente tenga tos y fiebre? $P(\mathit{Tos} = 1, \mathit{Fiebre} = 1)$ ó $P(\mathit{tos}, \mathit{fiebre})$
- ¿Cuál es la probabilidad de que un paciente tenga tos junto con cada valor posible de fiebre? $P(\mathit{Tos} = 1, \mathit{Fiebre})$ ó $P(\mathit{tos}, \mathit{Fiebre})$
- ¿Cuál es la probabilidad de que un paciente tenga/no tenga tos junto con cada valor posible de fiebre? $P(\mathit{Tos}, \mathit{Fiebre})$
- ¿Cuál es la probabilidad de que un paciente tenga/no tenga tos si se observa si tiene o no fiebre? $P(\mathit{Tos}|\mathit{Fiebre})$

Ejemplo: Tos y aroma picante

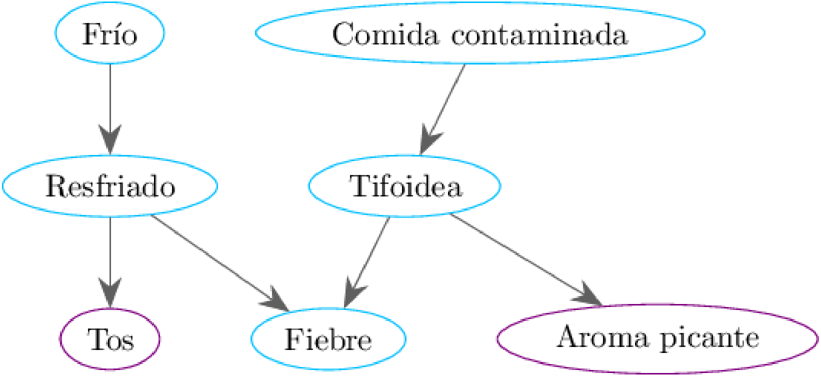


Figura: Consulta: $P(Tos, Aroma\ picante)$.

Método general

1. Identificar las variables que aparecen en la consulta.
2. Si la probabilidad a calcular es condicional, usar su definición para obtener por separado, numerador y denominador.
3. Determinar si hay rutas activas entre las variables participantes.
4. Identificar a los ancestros de las variables participantes.
5. Marginalizar a las variables que no aparecen explícitamente en la pregunta.
6. Usar la regla de la cadena para redes de Bayes.
7. Factorizar sumas y multiplicaciones lo mejor que se pueda.
8. Utilizar factores para resolver las multiplicaciones, marginalizaciones, reducciones probabilísticas necesarias.

Referencias I

Koller, Daphne y Nir Friedman (2009). *Probabilistic Graphical Models, Principles and Techniques*. MIT Press Cambridge.