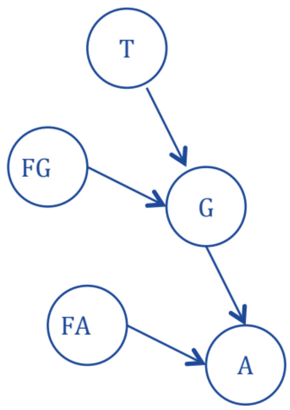
Inteligencia Artificial — Probl. de redes bayesianas

1. Redes bayesianas: central nuclear

En su central nuclear local, hay una alarma que detecta cuando un medidor de temperatura excede un umbral determinado. El medidor mide la temperatura del núcleo. Considera las variables booleanas A (suena la alarma), F_A (la alarma está defectuosa) y F_G (el medidor está defectuoso) y los nodos multivalor G (lectura del medidor) y T (temperatura real de núcleo).

(a) Dibuja una red bayesiana para este dominio, dado que es más probable que el medidor falle cuando la temperatura del núcleo es demasiado alta.



(b) Supón que sólo hay dos posibles temperaturas reales y medidas: normal y alta; la probabilidad de que el medidor dé la temperatura correcta es x cuando está funcionando, pero y cuando está defectuoso. Proporciona la tabla de probabilidad condicional asociada con G.

		Lectura Medidor (G)	
Temperatura Real (T)	**Medidor Defectuoso (FG)**	**Normal**	**Alta**
Normal	No	x	1-x
Normal	Sí	y	1-y
Alta	No	1-x	x
Alta	Sí	1-y	y

(c) Supón que la alarma funciona correctamente a menos que sea defectuosa, en cuyo caso nunca suena. Proporciona la tabla de probabilidad condicional asociada con A.

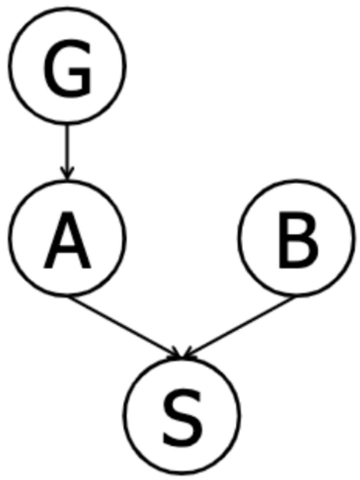
		Alarma suena (A)	
Lectura Medidor (G)	**Alarma defectuosa (FA)**	**Sí**	**No**
Normal	No	0	1
Normal	Sí	0	1
Alta	No	1	0
Alta	Sí	0	1

2. Probabilidad

Supongamos que un paciente puede tener un síntoma (S) que puede estar causado por dos enfermedades diferentes e independientes (A y B). Se sabe que la variación del gen G desempeña un papel importante en la manifestación de la enfermedad A. A continuación, se muestra un modelo y algunas tablas de probabilidad condicional para esta situación. Para cada parte, puede dejar su respuesta como una expresión aritmética.

$P(G)$	
$+g$	0.1
$-g$	0.9

$P(A\|G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9



$P(B)$	
$+b$	0.4
$-b$	0.6

$P(S\|A,B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

(a) Calcula la siguiente entrada de la distribución conjunta: $P(+g, +a, +b, +s)$

Para calcular la entrada de la distribución conjunta P(+g,+a,+b,+s), necesitamos considerar las dependencias condicionales y multiplicar las probabilidades relevantes de las tablas que tenemos en la red bayesiana.

La probabilidad conjunta P(+g,+a,+b,+s) se puede descomponer en un producto de probabilidades condicionales y marginales usando la regla de la cadena para redes bayesianas (diapositiva 19 del tema Redes bayesianas):

P(+g,+a,+b,+s) = P(+s|+a,+b) · P(+a|+g) · P(+b) · P(+g)

Podemos obtener cada uno de estos términos de las tablas de probabilidad condicional (CPT):

1. P(+g) es la probabilidad de que la variante genética +g esté presente.

2. P(+a|+g) es la probabilidad de que la enfermedad A esté presente dado que la variante genética +g está presente.

3. P(+b) es la probabilidad de que la enfermedad B esté presente.

4. P(+s|+a,+b) es la probabilidad de que el síntoma S esté presente dado que ambas enfermedades A y B están presentes.

La probabilidad conjunta $P(+g,+a,+b,+s)$ de que el paciente tenga la variante genética +g, la enfermedad A (+a), la enfermedad B (+b), y el síntoma S (+s), es de 0,04 o 4%.

(b) ¿Cuál es la probabilidad de que un paciente tenga la enfermedad A?

Para calcular la probabilidad de que un paciente tenga la enfermedad A, $P(+a)$, necesitamos sumar sobre todas las posibles variaciones del gen G, ya que la enfermedad A está condicionada al gen G. Utilizando las sumas totales y las probabilidades condicionales, el cálculo sería*:

$P(+a) = P(+a|+g)P(+g) + P(+a|-g)P(-g)$

Sustituyendo por valores de las tablas, la probabilidad de que un paciente tenga la enfermedad A, $P(+a)$, es de 0,19 o 19%.

* Tener en cuenta (diapositiva 10 del tema Probabilidad) que $P(A) = \sum_G P(A, G)$ y que por la regla del producto (diapositiva 14 del tema Probabilidad), $P(A, G) = P(A \mid G) \cdot P(G)$

(c) ¿Cuál es la probabilidad de que un paciente tenga la enfermedad A si tiene la enfermedad B?

Para calcular la probabilidad de que un paciente tenga la enfermedad A dado que ya tiene la enfermedad B, $P(+a|+b)$, necesitamos considerar si las enfermedades A y B son independientes. En una red bayesiana, la independencia se puede determinar por la estructura del grafo.

Si no hay una conexión directa entre A y B en el grafo, y no hay un nodo común que sea antecesor de ambos (como es el caso aquí, ya que G es antecesor de A pero no de B), entonces podemos asumir que A y B son independientes. Esto significa que el hecho de tener B no afecta la probabilidad de tener A.

Por lo tanto, la probabilidad de tener la enfermedad A dado que se tiene la enfermedad B es simplemente la probabilidad de tener la enfermedad A, que ya hemos calculado anteriormente como 19%.

$P(+a|+b) = P(+a) = 0{,}19$

Así que, bajo la suposición de independencia, la probabilidad de que un paciente tenga la enfermedad A dado que tiene la enfermedad B es de 0,19

(d) ¿Cuál es la probabilidad de que un paciente tenga la enfermedad A si tiene el síntoma S y la enfermedad B?

Para calcular la probabilidad de que un paciente tenga la enfermedad A dado que tiene el síntoma S y la enfermedad B, $P(+a|+s,+b)$, usamos la definición de probabilidad condicional. La fórmula relevante en este caso es (diapositiva 11 del tema Probabilidad):

$P(+a|+s,+b) = P(+a,+b,+s)/P(+b,+s) = P(+a,+b,+s)/P(+b,+s) = P(+a,+b,+s)/[P(+a,+b,+s) + P(-a,+b,+s)]$

Aquí, necesitamos calcular dos probabilidades conjuntas:

- $P(+a,+b,+s)$: La probabilidad de que el paciente tenga la enfermedad A, la enfermedad B y el síntoma S.

- $P(-a,+b,+s)$: La probabilidad de que el paciente no tenga la enfermedad A, pero tenga la enfermedad B y el síntoma S.

Estas probabilidades se calculan utilizando la regla de la cadena para descomponer las probabilidades conjuntas en un producto de probabilidades condicionales y marginales:

- $P(+a,+b,+s) = P(+a)P(+b)P(+s|+a,+b) = 0{,}19 \cdot 0{,}4 \cdot 1{,}0 = 0{,}076$

- $P(-a,+b,+s) = P(-a)P(+b)P(+s|-a,+b) = (1-0{,}19) \cdot 0{,}4 \cdot 0{,}8 = 0{,}2592$

Los valores necesarios para estos cálculos provienen de las tablas de probabilidad condicional del modelo.

Después de realizar estos cálculos, encontramos que:

- $P(+a,+b,+s)$ es aproximadamente 0,076.

- $P(-a,+b,+s)$ es aproximadamente 0,2592.

Insertando estos valores en la fórmula, obtenemos:

$\underline{P(+a|+s,+b)} = 0{,}076/(0{,}076 + 0{,}2592) = \underline{0{,}227}$

Por lo tanto, la probabilidad de que un paciente tenga la enfermedad A dado que tiene el síntoma S y la enfermedad B es aproximadamente 22.7%.

(e) ¿Cuál es la probabilidad de que un paciente porte la variación del gen G si tiene la enfermedad A?

Para calcular la probabilidad de que un paciente porte la variación del gen G dado que tiene la enfermedad A, $P(+g|+a)$, aplicaremos el teorema de Bayes. Esta probabilidad se calcula utilizando la información sobre cómo la presencia de la enfermedad A influye en nuestra creencia sobre la presencia de la variación genética G.

La fórmula del teorema de Bayes en este contexto es:

$P(+g|+a) = [P(+a|+g) \cdot P(+g)]/P(+a)$

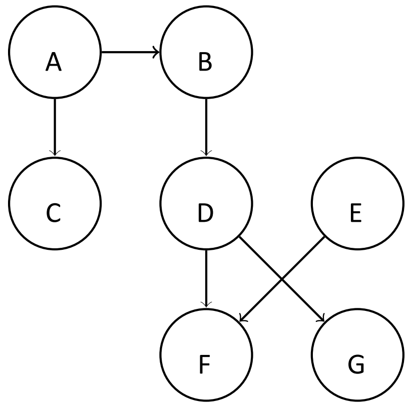
Ya hemos calculado $P(+a)$ como 0,19 en apartados anteriores. Los otros términos $P(+a|+g)$ y $P(+g)$ están dados por las tablas de probabilidad condicional del modelo. Vamos a proceder con este cálculo utilizando estos valores.

$\underline{P(+g|+a)} = (1{,}0 \cdot 0{,}1)/0{,}19 = \underline{0{,}526}$

La probabilidad de que un paciente porte la variación del gen G dado que tiene la enfermedad A, $P(+g|+a)$, es aproximadamente 0,526 o 52.6%.

3. Redes bayesianas: representación e independencia

Las partes (a) y (b) corresponden a la siguiente red bayesiana:



(a) Exprese la distribución de probabilidad conjunta como un producto de términos que representan tablas de probabilidades condicionales individuales asociadas a la Red de Bayes.

Para expresar la distribución de probabilidad conjunta de una red bayesiana, multiplicamos las probabilidades condicionales de cada nodo dado sus padres en la red. Para este gráfico, la distribución conjunta se expresaría generalmente de la siguiente manera:

Si tenemos una red con nodos A, B, C, D, E, F, y G, la distribución de probabilidad conjunta P(A, B, C, D, E, F, G) se expresaría como:

$P(A, B, C, D, E, F, G) = P(A) \cdot P(C|A) \cdot P(B|A) \cdot P(D|B) \cdot P(E) \cdot P(F|D, E) \cdot P(G|D)$

Cada término P(X|Y) representa la probabilidad condicional del nodo X dado el nodo o nodos Y que son sus padres en la red bayesiana. En el caso de nodos que no tienen padres (es decir, son causas raíz), simplemente usamos la probabilidad marginal P(X).

Para cada nodo, la tabla de probabilidades condicionales (CPT por sus siglas en inglés, Conditional Probability Table) proporciona las probabilidades de los posibles valores del nodo dado cada combinación de valores de sus nodos padres.

(b) Supongamos que cada nodo puede tomar 4 valores. ¿Cuántas entradas tienen los factores en A, D y F?

En una red bayesiana, el número de entradas en las tablas de probabilidad condicional (CPT) de cada nodo depende de cuántos valores puede tomar el nodo y cuántos valores pueden tomar sus padres.

- Para el nodo A: Como no tiene padres, su tabla de probabilidad condicional simplemente enumerará las probabilidades de cada uno de los 4 valores que puede tomar. Por lo tanto, habrá 4 entradas en la CPT de A.

- Para el nodo D: Dado que tiene un solo padre (B) y tanto B como D pueden tomar 4 valores cada uno, necesitamos una entrada para cada combinación posible de los valores de B y D. Esto significa que la CPT de D tendrá 4 (valores de B) x 4 (valores de D) = 16 entradas.

- Para el nodo F: F tiene dos padres (D y E), y cada uno puede tomar 4 valores. Por lo tanto, necesitaremos una entrada para cada combinación posible de los valores de D y E para cada valor que F puede tomar. Esto significa que la CPT de F tendrá 4 (valores de D) x 4 (valores de E) x 4 (valores de F) = 64 entradas.

En resumen, A tendrá 4 entradas, D tendrá 16 entradas y F tendrá 64 entradas en sus respectivas CPTs.

Considere las siguientes tablas de distribución de probabilidad. La distribución conjunta P(A,B,C,D) es igual al producto de estas tablas de distribución de probabilidad.

A	$P(A)$
+a	0.8
–a	0.2

A	B	$P(B\|A)$
+a	+b	0.9
+a	–b	0.1
–a	+b	0.6
–a	–b	0.4

B	C	$P(C\|B)$
+b	+c	0.8
+b	–c	0.2
–b	+c	0.8
–b	–c	0.2

C	D	$P(D\|C)$
+c	+d	0.25
+c	–d	0.75
–c	+d	0.5
–c	–d	0.5

Se están creando dispositivos de seguridad avanzados para coches que pueden avisar al conductor si se está quedando dormido (A) y también calcular la probabilidad de colisión (C) en tiempo real. Tienes a tu disposición 6 sensores (variables aleatorias):

- E: si el conductor tiene los ojos abiertos o cerrados

- W: si se toca o no el volante

- L: si el coche está en el carril o no

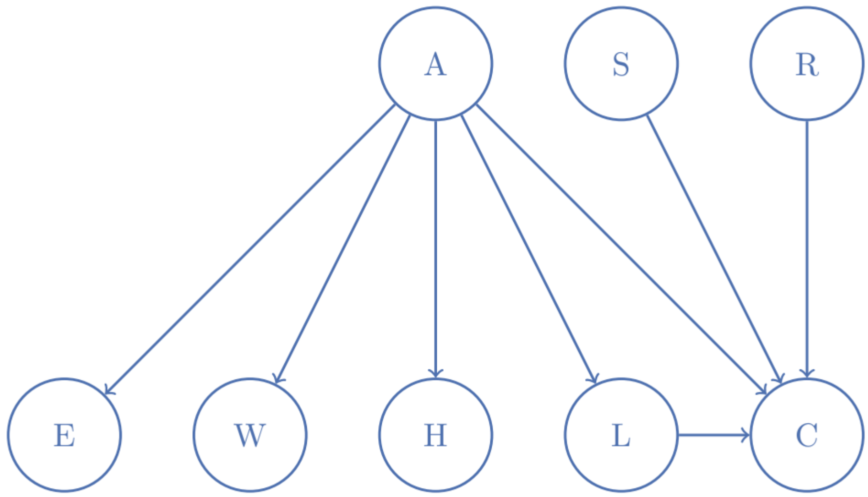
- S: si el coche va rápido o no

- H: si la frecuencia cardiaca del conductor es algo elevada o está en reposo

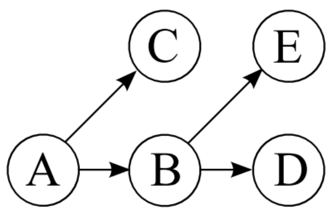
- R: si el radar del coche detecta un objeto cercano o no

A influye en $\{$E, W, H, L, C$\}$. C está influido por $\{$A, S, L, R$\}$.

(c) Dibuje la red bayesiana asociada a la descripción anterior añadiendo arcos entre los nodos proporcionados cuando proceda.


A
S
R
E
W
H
L
C

4. Eliminación de variables



	$P(A)$
$+a$	0.25
$-a$	0.75

$P(B\|A)$	$+b$	$-b$
$+a$	0.5	0.5
$-a$	0.25	0.75

$P(C\|A)$	$+c$	$-c$
$+a$	0.2	0.8
$-a$	0.6	0.4

$P(D\|B)$	$+d$	$-d$
$+b$	0.6	0.4
$-b$	0.8	0.2

$P(E\|B)$	$+e$	$-e$
$+b$	0.25	0.75
$-b$	0.1	0.9

(a) Utilizando la red de Bayes y las tablas de probabilidad condicional anteriores, calcula las siguientes cantidades:

(i) $P\,(+b\,|\,+a)$ = 0,5

Para calcular la probabilidad condicional P(+b | +a), necesitamos referirnos a la tabla de probabilidad condicional de B dado A, P(B|A), porque nos interesa saber la probabilidad de que B sea verdadero dado que A es verdadero.

(ii) $P\,(+a,+b)$ = 0,25 · 0,5 = 0,125 = 1/8

Para calcular la probabilidad conjunta P(+a, +b), necesitamos utilizar la regla del producto de las probabilidades en la red bayesiana, que es:

P(+a, +b) = P(+b | +a) · P(+a)

Ya hemos identificado P(+b | +a) = 0.5 de la tabla de probabilidad condicional P(B|A).

La probabilidad marginal P(+a) se puede obtener directamente de la tabla de probabilidad marginal de A, la cual, según la imagen, es 0,25.

Entonces, aplicando la regla del producto:

P(+a, +b) = 0.5 · 0.25

(iii) $P\,(+a\,|\,+b)$ = (0,25 · 0,5) / (0,25 · 0,5 + 0,25 · 0,75) = 0,4 = 2/5

Para calcular P(+a|+b), la probabilidad de que A sea verdadero dado que B es verdadero, podemos usar la regla de Bayes, que se define como:

P(+a|+b) = [P(+b|+a) · P(+a)]/P(+b)

Ya tenemos P(+b|+a) = 0,5 y P(+a) = 0,25, pero necesitamos P(+b), la probabilidad marginal de B, para completar el cálculo. P(+b) no se proporciona directamente en las tablas, por lo que debe calcularse sumando las probabilidades conjuntas de B y todos las posibles estados de A, que son:

P(+b) = P(+b|+a) · P(+a) + P(+b|-a) · P(-a)

Donde:

- P(+b|+a) es la probabilidad de +b dado +a, que es 0,5.
- P(+a) es la probabilidad de +a, que es 0,25.
- P(+b|-a) es la probabilidad de +b dado -a, que debemos buscar en la tabla de probabilidad condicional P(B|A). Es 0,25.
- P(-a) es la probabilidad de -a, que es el complemento de P(+a), es decir, 1 - P(+a) = 0,75

Una vez que tengamos P(+b), podemos calcular P(+a|+b) utilizando la regla de Bayes.

(b) Ahora vamos a considerar la eliminación de variables en la red de Bayes anterior

(i) Supongamos que tenemos la evidencia $+c$ y deseamos calcular $P\ (E \mid +c)$. ¿Qué factores tenemos inicialmente?

La eliminación de variables es un proceso en el razonamiento probabilístico con redes bayesianas, que se utiliza para calcular la probabilidad posterior de un conjunto de variables de interés, dado algún conjunto de evidencias. Este proceso implica sumar (marginalizar) sobre las variables que no son de interés.

Dado que tenemos la evidencia +c, estamos interesados en calcular P(E | +c). Para hacerlo, primero identificamos los factores iniciales que están presentes en nuestra red bayesiana. Un factor es básicamente una función que representa un conjunto de probabilidades condicionales o marginales que están relacionadas a través de las dependencias de la red.

En la red bayesiana proporcionada, los factores iniciales considerando la evidencia +c serían:

$P\ (A), P\ (B \mid A), P\ (+c \mid A), P\ (D \mid B), P\ (E \mid B)$

(ii) Si eliminamos la variable B, creamos un nuevo factor. ¿Cuál es la ecuación para calcular el factor que creamos al eliminar la variable B?

$f(A, D, E) = \sum_B P\ (B \mid A) \cdot P\ (D \mid B) \cdot P\ (E \mid B)$

(iii) Tras eliminar la variable B, ¿cuál es el nuevo conjunto de factores? Para cada factor, indique también su tamaño.

Factor	Tamaño después de la eliminación
P(A)	2
P(+c | A)	2
$f(A, D, E)$	2^3

(iv) Supongamos ahora que tenemos la prueba $-c$ y estamos tratando de calcular $P\ (A\,|\,-\,c)$. ¿Cuál es el orden de eliminación más eficiente? Si más de una ordenación es la más eficiente, proporcione cualquiera de ellas

Para calcular P(A | -c) con evidencia de -c, queremos eliminar las variables que no son A o C de la forma más eficiente posible. En la eliminación de variables, el objetivo es reducir la complejidad computacional al minimizar el tamaño de los factores intermedios que se generan durante el proceso. Generalmente, queremos eliminar primero las variables que tienen la menor cantidad de dependencias (es decir, están menos conectadas en la red).

La red bayesiana dada tiene la estructura A → C, A → B → D, y B → E. Dado que E no tiene padres aparte de B y no es antecesor de ninguna otra variable en la red, es la variable más independiente. Así que deberíamos empezar eliminando E. Después de E, D solo depende de B, por lo que D debería eliminarse a continuación. B está conectado a A, y es necesario para calcular la probabilidad de C, por lo que deberíamos mantener B hasta que A y C se hayan considerado. A afecta tanto a B como a C y no puede ser eliminado hasta que estas otras variables sean consideradas.

Entonces, un orden eficiente de eliminación de variables para calcular P(A | -c) sería:

1. Eliminar E (porque solo depende de B y no está conectado a C o A).

2. Eliminar D (porque una vez que E ha sido eliminado, D solo depende de B).

3. Eliminar B (ya que ahora hemos eliminado D y E que dependían de B, y solo nos quedan A y C).

4. Finalmente, usaríamos el factor resultante que involucra a A y C junto con la evidencia -c para calcular P(A | -c).

Este orden, E, D, B, minimiza el tamaño de los factores intermedios porque eliminamos primero las variables que tienen menos conexiones, lo que simplifica los cálculos en cada paso.

Se podría hacer un razonamiento equivalente con D, E, B, que también es un ordenamiento válido.

(v) Una vez que hemos ejecutado la eliminación de variables y tenemos $f(A,\text{-}c)$ ¿cómo calculamos $P\ (+a\,|\,-\,c)$?

Una vez que hemos ejecutado la eliminación de variables, nos queda un factor f(A, -c) que representa la distribución de probabilidad no normalizada de A dado que hemos observado -c. Para calcular la probabilidad condicional P(+a | -c), necesitamos normalizar este factor. La normalización implica asegurarse de que las probabilidades sumen 1, convirtiendo la distribución no normalizada en una distribución de probabilidad válida.

El proceso es el siguiente:

1. Calcular el factor no normalizado para +a y -a:

Dado que ya hemos eliminado las demás variables, f(A, -c) nos da los valores para +a, -c y -a, -c (es decir, f(+a, -c) y f(-a, -c)).

2. Sumar ambos factores para obtener la constante de normalización:

Esta suma será la constante de normalización Z, que es la suma de todos los valores del factor para A dado -c.

Z = f(+a, -c) + f(-a, -c)

3. Dividir el factor no normalizado por la constante de normalización para obtener la probabilidad condicional:

La probabilidad condicional normalizada de +a dado -c se obtiene dividiendo el factor no normalizado por Z.

P(+a | -c) = f(+a, -c) / Z

Si f(A, -c) ya tiene los valores para f(+a, -c) y f(-a, -c), solo necesitamos ejecutar estos pasos para obtener la probabilidad condicional deseada.

En resumen:

$$\frac{f(+a,-c)}{f(+a,-c)+f(-a,-c)}$$