

INTELIGENCIA ARTIFICIAL

Práctica 4

Curso 2018-2019

Autores:

Rubén Rodríguez Esteban 737215

Ejercicio 1

Consideremos esta red Bayesiana para diagnóstico de enfermedades del pulmón. Se nos pide responder razonadamente a las siguientes cuestiones:

Apartado A

- Son A y B condicionalmente independientes dado X

A) A y B son independientes dado X

Seleccionamos los caminos entre A y B marcando X en verde

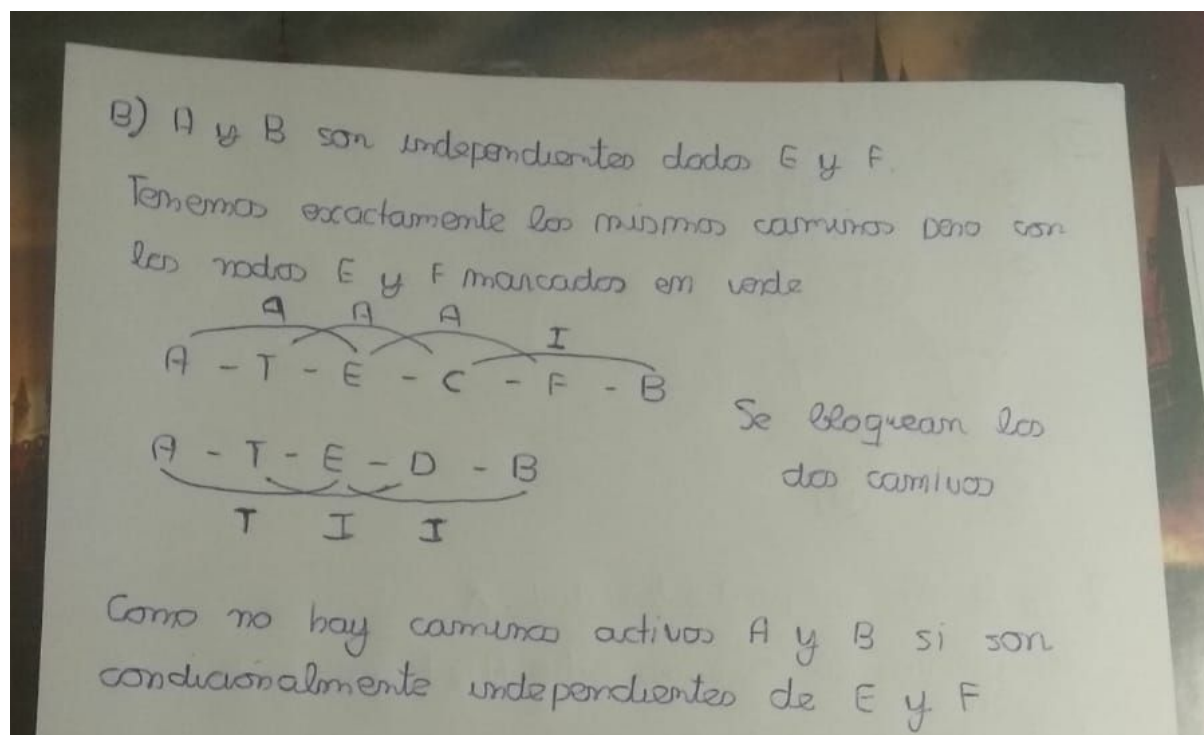
$A - T - E - C - F - B$

$A - T - E - D - B$

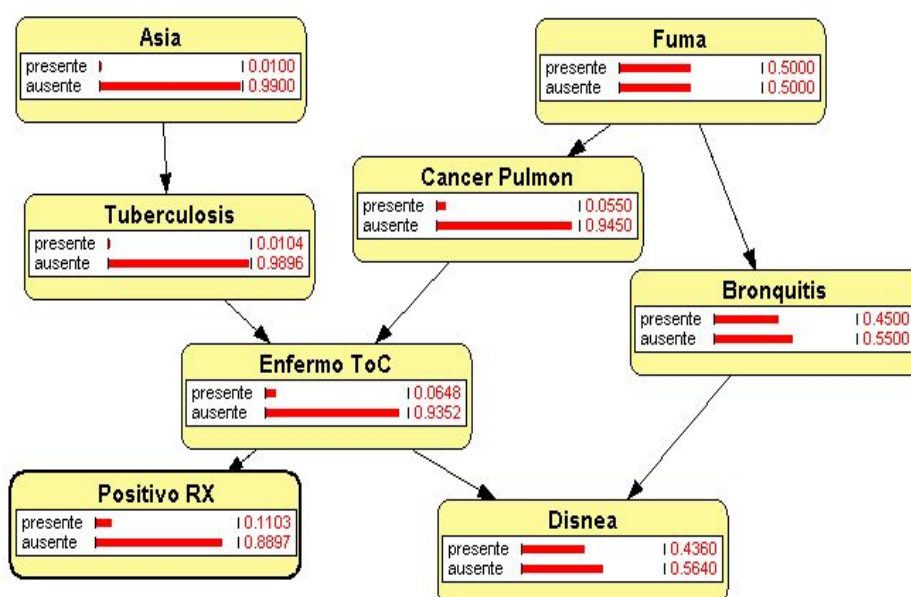
En el primer caso o camino son todos los tripletes activos por lo que es un camino activo. El segundo camino es bloqueado porque el triplete E-D-B no es activo.

Como hay un camino activo no son condicionalmente independientes

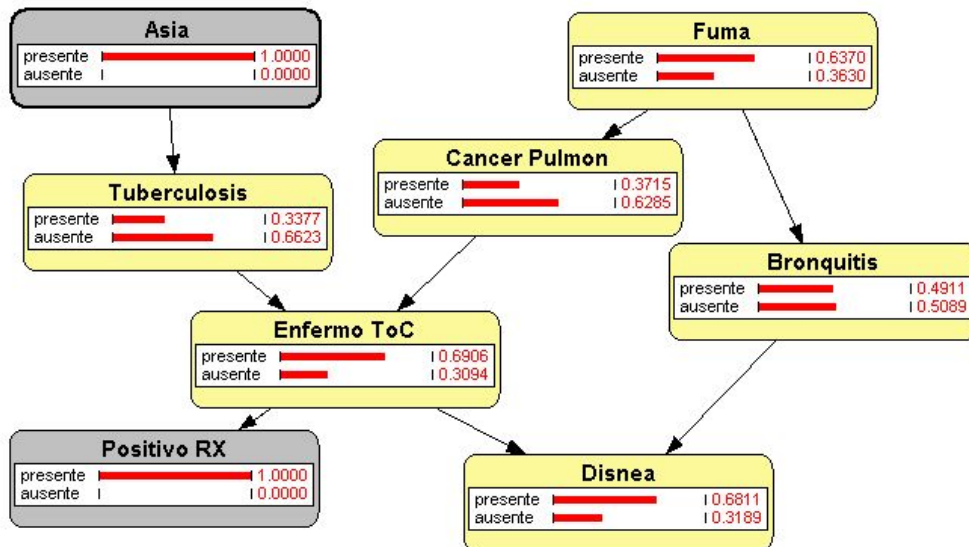
- Son A y B condicionalmente independientes dado E y F



En la red generada con el programa openmarkov se pueden observar los resultados a los que se ha llegado

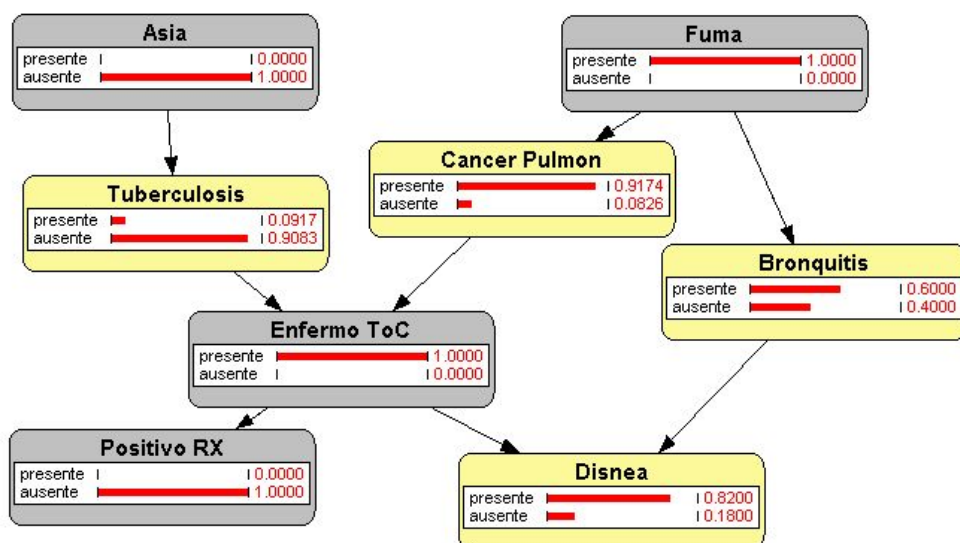
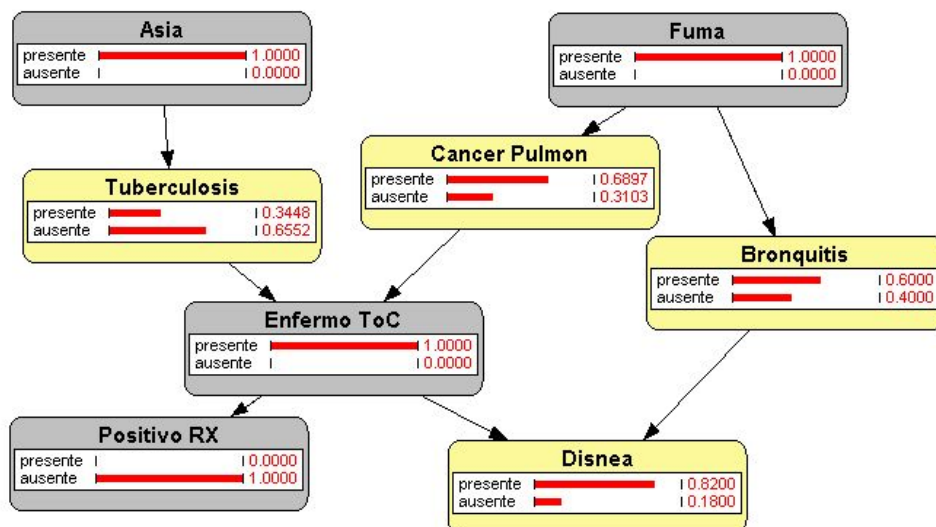


Podemos observar que tal y como se afirmó en el trabajo previo de este apartado, efectivamente se puede probar que si son condicionalmente independientes de X. Para poderlo probar se ha procedido a dejar la X presente junto con la A tal y como indica la imagen inferior, y posteriormente se ha dejado en modo ausente.



En estas dos imágenes se puede observar como A y B si que dependen de la variable X ya que ya se puede apreciar como en la variable B se modifican los valores dejando fijada la la X y cambiando la A de presente a ausente y viceversa. Por consiguiente A y B dependen de de la variable X.

De la misma forma se puede observar que las variables A y B no que son independientes de ya que al poner las variables E y F como presentes, si se realizan modificaciones sobre la variable A, la B no registra ningún cambio y viceversa, tal y como evidencian las imágenes. Por consiguiente se puede afirmar que los resultados obtenidos en el previo son correctos.



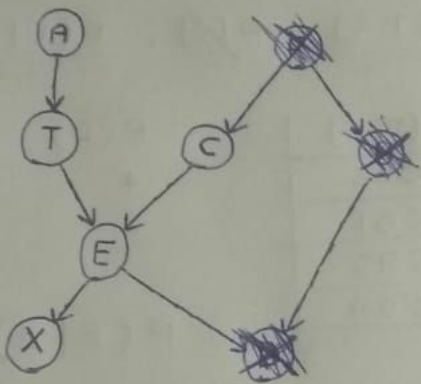
Apartado B

Calcular siguiendo el algoritmo de eliminación de variables la probabilidad de que tenga tuberculosis un paciente que ha dado positivo en rayos X y fuma, si se sabe que no tiene cáncer de pulmón. Si hay alguna variable que se pueda ignorar, explica claramente por qué.

En las capturas siguientes se muestran tanto el procedimiento como los resultados obtenidos en el trabajo previo de este apartado.

b)

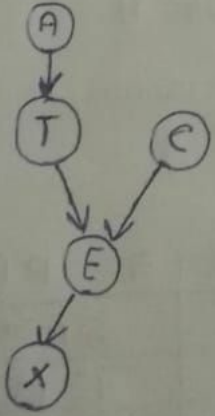

$P(T|X, F, \bar{C}) \ni$ probabilidad que se debe calcular



1 \ni Ignoración de variables

Ignoramos B y D porque no son ancestros de una Query o evidencia

Eliminamos la variable F por el triplete activo



\ni esta es la red bayesiana resultante

2. Instanciar las tablas de probabilidad con la evidencia

$P(A)$, $P(T|A)$, $P(E|T,C)$, $P(X)$, $P(E|X)$

$$P(A)$$

+	0.01
-	0.99

T	A	$P(T A)$
+	+	0.05
+	-	0.01
-	+	0.95
-	-	0.99

$P(E|X)$

E	T	C	$P(E T,C)$
+	+	-	1
+	-	-	0
-	+	-	0
-	-	-	1

X	E	$P(X E)$
+	+	0.98
+	-	0.05

3. Eliminación para cada variable oculta H

- Sumar todos los factores que mencionan H
- Eliminar H sumando

$P(T A)$	$P(A)$	$P(T A) \times P(A)$	\exists	$P(T, A)$
+	+	0.05	0.01	$5 \cdot 10^{-4}$
+	-	0.095	0.01	$9.5 \cdot 10^{-3}$
-	+	0.95	0.99	$9.9 \cdot 10^{-3}$
-	-	0.99	0.99	0.9801

$P(T)$		
+	$5 \cdot 10^{-4}$	$+ 0.0099 = 0.0104$
-	0.0095	$+ 0.9801 = 0.9896$

Eliminación de la variable E

$P(E T,C)$			$P(E X)$		X	$P(E,T,C,X)$
+	+	-	+	+	$1 \cdot 0.98$	0.98
+	-	-	+	+	$0 \cdot 0.98$	0
-	+	-	-	+	$0 \cdot 0.005$	0
-	-	-	-	+	$1 \cdot 0.005$	0.005

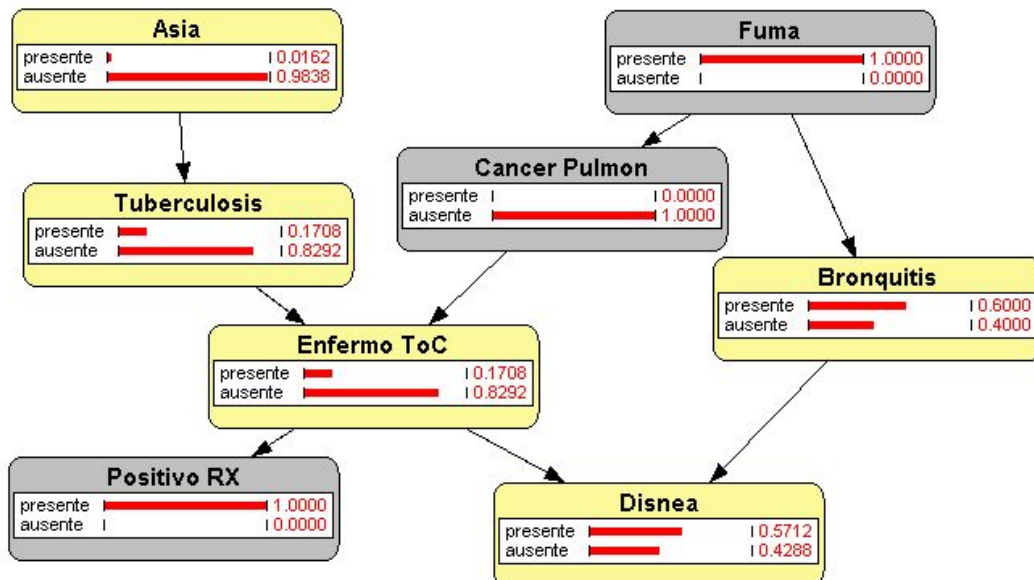
$P(T,-C,X)$	$P(T,-C,X)$
0.98	$+t -c +x$
0.005	$-t -c +x$

sumamos con
respecto de
E

$P(T,-C,X)$	$P(T)$	$P(T)$
$+t -c +x$	t	0.010192
$-t -c +x$	-t	0.04948
		0.059672

$\Rightarrow 0.1708 //$
 $\Rightarrow 0.8291 //$

Con la imagen siguiente de openmarkov se puede comprobar que los cálculos realizados son correctos:



Ejercicio 2

En un estudio sobre discriminación de género en la admisión de la Universidad de Berkeley, California, se recogieron los siguientes datos:

Carrera	Hombres		Mujeres	
	Solicitudes	Admitidos	Solicitudes	Admitidas
A	825	512	108	89
B	560	353	25	17
C	325	120	593	202
D	417	138	375	131

En base a estos datos se pide responder a las cuestiones siguientes:
Apartado A

Representa el problema con una Red Bayesiana causal. Piensa muy bien cuáles son las variables aleatorias necesarias, y cuáles las relaciones que pueden aparecer entre ellas. No hagas ninguna hipótesis de independencia entre variables salvo que esté soportada por los datos de la tabla.

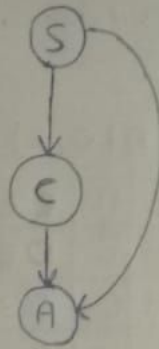
2

a)

$S \Rightarrow \text{sexo}$

$C \Rightarrow \text{Tipo de cama}$

$A \Rightarrow \text{admisión}$



En la red bayesiana no se ha supuesto ninguna hipótesis adicional. $S \rightarrow A$ significa que el sexo es influyente en la admisión a la cama.

b)

$P(S)$

Se calcula así.

$$\text{Personas} = 825 + 560 + 315 + 417 + 108 + 25 + 313 + 375 = \underline{3228} \text{ personas}$$

$$H \Rightarrow \frac{2127}{3228} \Rightarrow 0.6589$$

$$M \Rightarrow \frac{1101}{3228} \Rightarrow 0.34107$$

$P(S)$

+S	0.6589
-S	0.34107

Apartado B

Escribe las tablas de probabilidad de cada nodo, estimando las probabilidades a partir de los datos de la tabla.

$$P(C|S) \Rightarrow \frac{\text{Solicitudes por carrera}}{\text{total solicitudes}} \Rightarrow \frac{\text{seco}}{\text{seco}}$$

$$P(S)$$

H	0.6589
M	0.34107

$$P(C|S)$$

H	A	0.3878
H	B	0.2632
H	C	0.1579
H	D	0.19605
M	A	0.098
M	B	0.0277
M	C	0.5386
M	D	0.3405

$$P(A|C, S)$$

H	A	0.6206
H	B	0.6303
H	C	0.3692
H	D	0.3309
M	A	0.82407
M	B	0.68
M	C	0.3406
M	D	0.3493

$$\Rightarrow \frac{\text{Admitidos}}{\text{Solicitudes}}$$

c) Se puede observar que si existe una discriminación muy clara en la admisión en favor de las mujeres en la carrera A ya que el hombre tiene 0.62 y la mujer 0.82, mientras que en el resto de carreras si bien es cierto que puede haber y hay diferencia, no es tan latente.

Apartado C

A la vista de las tablas de probabilidad, ¿ha podido haber discriminación de género en la admisión?
Razona la respuesta.

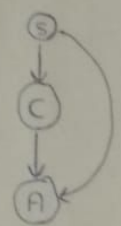
Se puede observar en la imagen

Apartado D

Utilizando el algoritmo de eliminación de variables, calcula la probabilidad de admisión para una mujer.

D) Calcular $P(A | S = m)$

1. Ignoramos aquellos que no son de Query o evidencia, y posibles tripletes activos



No se puede eliminar nodo

2. Instanciamos las tablas con la evidencia $S = m$

$P(S)$

m	0.34107
---	---------

la evidencia es $S = m$

$P(C | S)$

m	A	0.098
m	B	0.0277
m	C	0.5386
m	D	0.3405

$P(A | C, S)$

m	A	0.82407
m	B	0.68
m	C	0.3406
m	D	0.3493

3. Para cada variable oculta juntamos factores y sumamos con respecto de ella

$P(A | C, S) \times P(C | S) = P(A, C, S) \quad P(A, S)$

m	A	0.82407
m	B	0.68
m	C	0.3406
m	D	0.3493

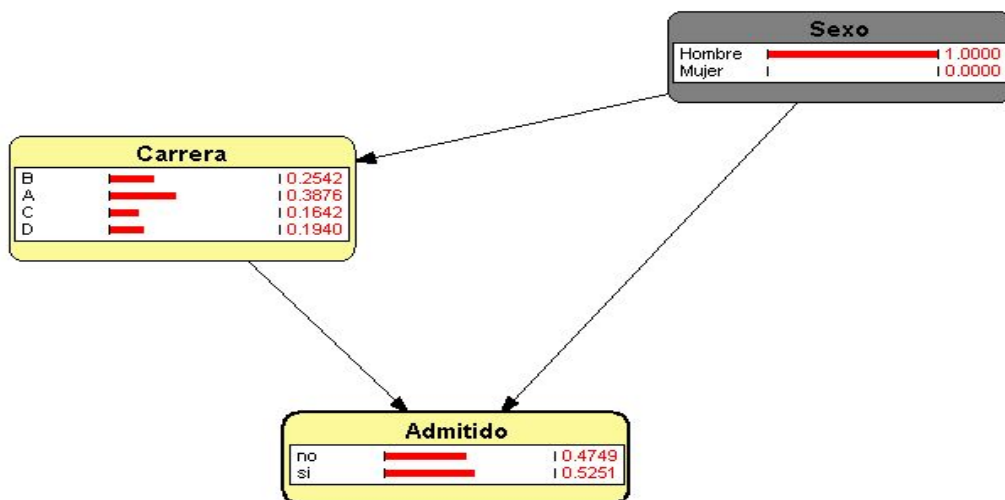
m	A	0.098
m	B	0.0277
m	C	0.5386
m	D	0.3405

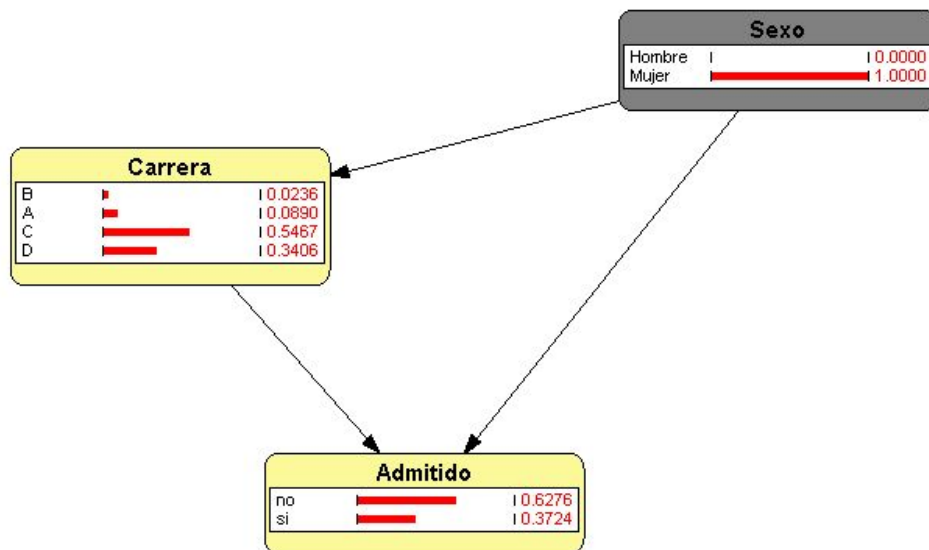
\Rightarrow

m	A	0.08075
m	B	0.015436
m	C	0.18344
m	D	0.1189

\Rightarrow 0.398526

Seguidamente en la siguiente captura se puede observar que los resultados obtenidos en el trabajo previo son los correctos tanto para la probabilidad de admisión del hombre como de la mujer:



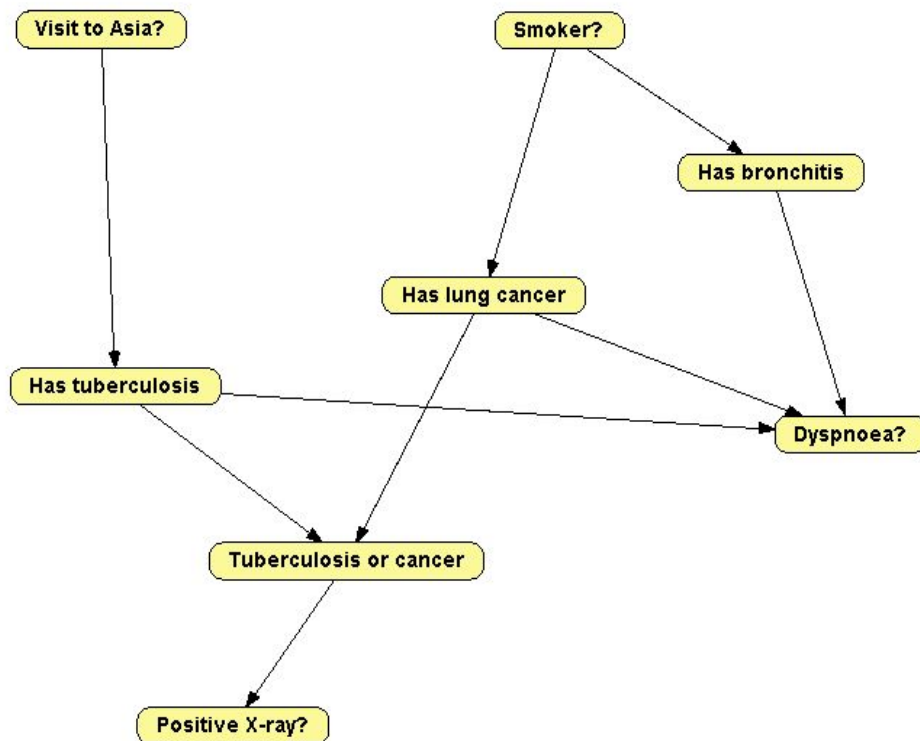


Ejercicio 3

El fichero DatosAsia.xls contiene datos simulados del ejemplo de enfermedades del pulmón. Prueba a aprender la red a partir de estos datos, con la opción de aprendizaje automático

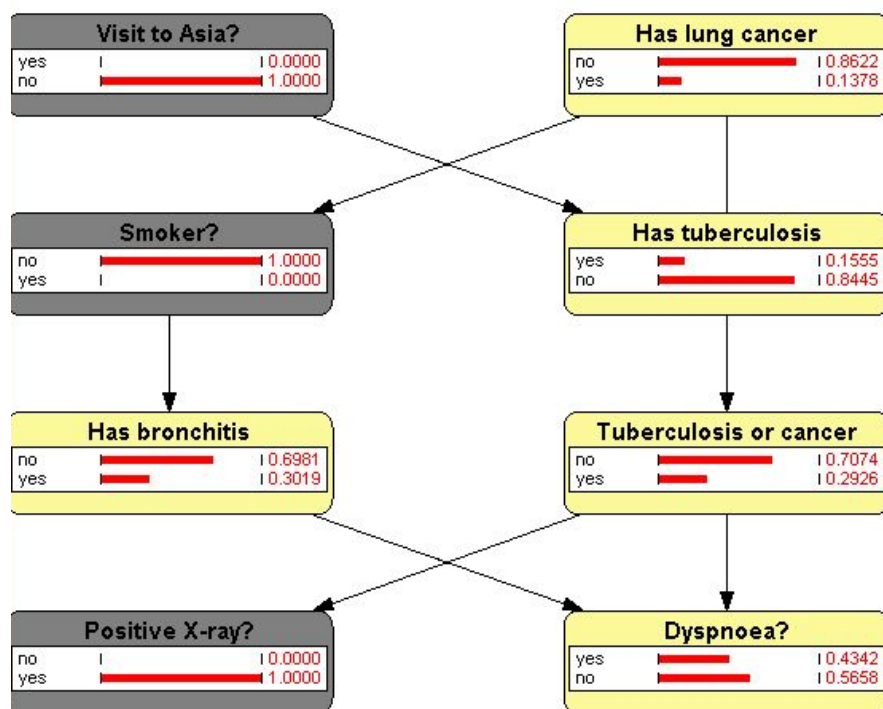
Apartado A

Si se que se puede observar que es una relación causal dado que existen relaciones causales como por ejemplo el factor de fumar y de contraer cáncer de pulmón, o el hecho de fumar y tener disnea. También es muy notable la relación causal en que si tienes cáncer de pulmón, disnea o tuberculosis estás enfermo y viceversa. A continuación se muestra una foto de la red causal resultante del fichero



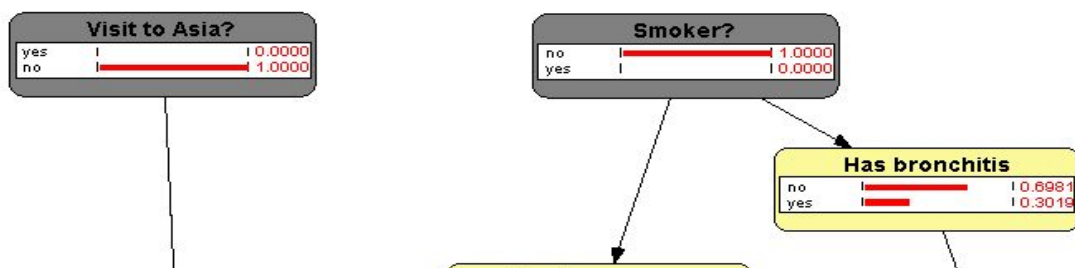
Apartado B

Con la red obtenida, calcula la probabilidad de que tenga tuberculosis un paciente que ha dado positivo en rayos X, no fuma, y no ha visitado Asia. Haz una captura de pantalla con la red para el informe.



Apartado C

Repite el proceso con la opción de aprendizaje interactivo, forzando manualmente a que aparezcan sólo enlaces causales.



Apartado D

Calcula con la nueva red la probabilidad de que tenga tuberculosis un paciente que ha dado positivo en rayos X, no fuma, y no ha visitado Asia. ¿Sale lo mismo que en el apartado b)? Analiza y discute por qué. Incluye captura de pantalla en el informe.

No sale lo mismo que en el apartado b debido a que no se tienen