Modelos Gráficos Probabilísticos

13 de abril de 2015

1. Se sabe que una prueba para la detección de una cierta enfermedad da positiva en el 96 % de los casos en que se está enfermo, y negativa en el 94 % de los sanos. Cierta persona se somete a la prueba y se sabe que, a su edad, una de cada 145 personas está enferma sin saberlo.

   a) ¿Cuál es la probabilidad de que la prueba de positiva?

   b) Si el resultado es positivo, ¿cuál es la probabilidad de que esté enferma realmente?

   c) Si la prueba fuese negativa, ¿cuál es la probabilidad de que a pesar de todo esté enferma?

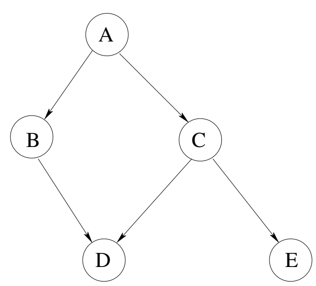
2. Da ejemplos prácticos de tres variables X, Y, Z en las que se cumpla:

   a) X es independiente de Y, pero X es dependiente de Y dado Z.

   b) X es dependiente de Y, pero X es independiente de Y dado Z.

   Dibuja las redes bayesianas que representen las relaciones entre las variables de estos ejemplos.

3. Supongamos la siguiente red bayesiana donde todas las variables tienen dos estados (variables binarias).



   Supongamos que tenemos los siguientes valores de probabilidad:

   - $P(a) = 0{,}2$
   - $P(b|a) = 0{,}8$ y $P(b|\overline{a}) = 0{,}2$
   - $P(c|a) = 0{,}2$ y $P(c|\overline{a}) = 0{,}05$
   - $P(d|b, c) = 0{,}9$ y $P(d|b, \overline{c}) = 0{,}9$
   - $P(d|\overline{b}, c) = 0{,}7$ y $P(d|\overline{b}, \overline{c}) = 0{,}05$
   - $P(e|c) = 0{,}8$ y $P(e|\overline{c}) = 0{,}6$

   a) Escribe la tabla de probabilidad para los potenciales asociados a las distribuciones $P(D|B, C)$ y $P(C|A)$.

   b) Escribe un árbol de probabilidad para el potencial de $P(D|B, C)$.

   c) Calcula el resultado de la combinación de los potenciales $P(D|B, C)$ y $P(C|A)$ representados con tablas.

   d) Marginaliza sobre las variables $\{A, B, C\}$ (o sea, borra la variable D) el potencial obtenido como resultado en el punto anterior.

4. Supongamos la red bayesiana del ejercicio anterior.

   a) Usando el algoritmo de eliminación de variables, queremos obtener la distribución de probabilidad a posteriori para la variable D considerando que no tenemos ninguna observación. Suponiendo que las variables se eliminan siguiendo el orden E, C, B, A, detalla paso a paso los cálculos que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo.

   b) Para el algoritmo Shafer-Shenoy, construir un árbol de grupos asociado a la red bayesiana anterior siguiendo el orden de eliminación E, D, C, B, A.

5. Describe brevemente las características de las dos grandes clases de métodos para el aprendizaje de la estructura de una red bayesiana.

6. Se desea estimar la probabilidad condicionada de una variable Y dado su conjunto de variables padres X_1, X_2,...,X_m en una red bayesiana, a partir de una base de datos (completa) que contiene instancias de todas las variables presentes en la red. Describe cuál sería el estimador escogido y por qué.

7. Dado el conjunto de variables $X = X_1, X_2, X_3, X_4, C$ todas con tres casos,

   a) Dibuja una red para cada una de las siguientes topologías: BAN, augmented naive bayesian k-dependiente con k=2, NB y TAN.

   b) Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red.

   c) Ordena los modelos de menor a mayor complejidad.

8. Cuál es la topología de una red obtenida por algún algoritmo de clasificación de los denominados Semi, como el Semi-Naive de Kononenko o Semi de Pazzani 2002

Modelos Gráficos Probabilísticos

17 de marzo de 2016

1. Tenemos dos urnas, una de ellas tiene 6 bolas blancas y 4 rojas y la segunda tiene 2 bolas blancas y 8 rojas. Consideremos el experimento consistente en tirar una moneda y si sale cara elegimos la primera urna y si sale cruz la segunda urna. Entonces, de la urna elegida extraemos una bola al azar.

   - Calcular la probabilidad de que la bola elegida sea blanca.
   - Calcular la probabilidad de que se haya elegido la urna primera si se sabe que la bola ha sido blanca.

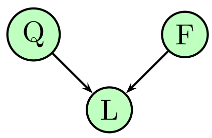
   Nota: se supone que la moneda está bien equilibrada y que la probabilidad de cara es la misma que la de cruz.

2. Consideremos que en una población mido las siguientes variables:

   - X Edad
   - Y Sexo
   - Z Cáncer de próstata
   - T Nivel de PSA en sangre (un marcador del cáncer de próstata)

   Dibujar un grafo dirigido acíclico que represente correctamente las independencias condicionadas existentes entre las variables del problema. Describir las independencias básicas representadas.

3. Supongamos la siguiente red bayesiana donde todas las variables tienen dos estados (variables binarias).

   Supongamos que tenemos los siguientes valores de probabilidad:

   - $P(q) = 0{,}7$
   - $P(f) = 0{,}6$
   - $P(l|q, f) = 0{,}98$ y $P(l|q, \overline{f}) = 0{,}8$
   - $P(l|\overline{q}, f) = 0{,}9$ y $P(l|\overline{q}, \overline{f}) = 0{,}0$

   a) Escribe la tabla de probabilidad para el potencial asociado a las distribuciones $P(L|Q, B)$ y $P(Q)$.

b) Calcula el resultado de la combinación de los potenciales $P(L|Q, B)$ y $P(Q)$ representados con tablas.

c) Marginaliza sobre las variables $\{L, F\}$ (o sea, borra la variable Q) el potencial obtenido como resultado en el punto anterior.

4. Supongamos la red bayesiana del ejercicio anterior.

   a) Usando el algoritmo de eliminación de variables, queremos obtener la distribución de probabilidad a posteriori para la variable L considerando que no tenemos ninguna observación. Suponiendo que las variables se eliminan siguiendo el orden Q, F detalla paso a paso los cálculos que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo.

   b) Cuando tenemos observadas algunas de las variables de una red bayesiana, ¿cuál método aproximado funciona mejor y por qué?, ¿muestreo lógico (*logic sampling*) o ponderación por verosimilitud (*likelihood weighting*)?

5. Describe dos métodos alternativos para estimar la probabilidad condicionada de una variable Y dado su conjunto de variables padres $X_1, X_2, \ldots, X_m$ en una red bayesiana, a partir de una base de datos (completa) que contiene instancias de todas las variables presentes en la red. Comenta las ventajas e inconvenientes de cada método. ¿Qué se puede hacer si la base de datos no es completa?

6. Describe con cierto detalle las características de los métodos de aprendizaje de la estructura de una red bayesiana basados en funciones de evaluación y técnicas de búsqueda. Pon ejemplos de los distintos elementos que configuran este tipo de métodos.

7. Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, C\}$ todas con tres casos,

   - dibuja una red para cada una de las siguientes topologías: BAN, augmented naive bayesian k-dependiente con k=2, NB y TAN.
   - Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red
   - Ordena los modelos de menor a mayor complejidad.

8. Indica alguna de las ventajas del método AODE (Averaged One DEpendence)

Modelos Gráficos Probabilísticos
20 de marzo de 2017

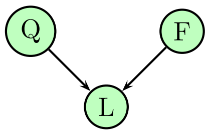
1. Hay una enfermedad que se da en el 1 % de la población. Se tienen dos tests para detectarla. El primero de ellos da positivo en el 90 % de los que tienen la enfermedad y en el 15 % de los que no la tienen. El segundo de ellos da positivo en el 85 % de los que tienen la enfermedad y en el 5 % de los que no la tienen. Se supone que los resultados de los tests son condicionalmente independientes dado que se conoce si se tiene o no se tiene la enfermedad. Calcular la probabilidad de que una persona seleccionada al azar tenga la enfermedad después de hacerle los tests y resultar los dos positivos.

2. Ha habido un crimen para el que hay un sospechoso y se pretende analizar la evidencia con una red bayesiana. Se prentende analizar la relación previa del sospechoso con la víctima, una mancha de sangre encontrada en la escena y una huella de zapato. Las variables que vamos a considerar son:

   - R: el sospechoso tenía una relación sentimental con la víctima (con valores Si y No)
   - H: el sospechoso es culpable
   - G: la mancha de sangre es del criminal
   - T: la huella es del crimimal
   - S: la mancha de sangre es del sospechoso
   - U: la huella es del sospechoso
   - E: El análisis de DNA de la mancha coincide con el sospechoso
   - F: La huella del zapato es compatible con el sospechoso

   Dibujar un grafo dirigido acíclico que represente correctamente las independencias condicionadas existentes entre las variables del problema. Hacer explícitas hipétesis de independencia que se hayan usado y que resulten dudosas en el enunciado del problema.

3. Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos (variables binarias).



   Supongamos que tenemos los siguientes valores de probabilidad:

   - $P(q) = 0{,}7$
   - $P(f) = 0{,}6$
   - $P(l|q, f) = 0{,}98$ y $P(l|q, \overline{f}) = 0{,}8$
   - $P(l|\overline{q}, f) = 0{,}9$ y $P(l|\overline{q}, \overline{f}) = 0{,}0$

   a) Escribe las tablas de probabilidad para los potenciales asociados a las distribuciones $P(Q)$, $P(F)$ y $P(L|Q, F)$.

   b) Escribe la tabla de probabilidad que se obtendría para el potencial asociado a la distribución $P(L|Q, F)$ si consideramos que viene definido con un *modelo noisy-or* donde tenemos los siguientes parámetros: probabilidad 0,8 de que Q cause L, probabilidad 0,9 de que F cause L y una probabilidad residual de $\lambda_0 = 0{,}0001$. Los números de la tabla de probabilidad pueden dejarse indicados como el producto de los números necesarios en cada caso.

4. Supongamos la red bayesiana del ejercicio anterior (apartado a).

   a) Usando el algoritmo de eliminación de variables, queremos obtener la distribución de probabilidad a posteriori para la variable L considerando que no tenemos ninguna observación. Suponiendo que las variables se eliminan siguiendo el orden $\{Q, F\}$, detalla paso a paso los cálculos que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo. Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total.

   b) Supongamos que aplicamos el algoritmo de muestreo lógico (*logic sampling*) para obtener la distribución de probabilidad a posteriori para la variable L considerando que tenemos observada la variable Q con el valor $Q = q$. Muestreamos las variables de la red siguiendo el orden Q, F y L. Indica cual es la configuración obtenida para las tres variables $\{Q, F, L\}$ en la primera realización, suponiendo que la realización se obtuvo al generar los siguientes números de una distribución uniforme: $\{0{,}1, 0{,}7, 0{,}6\}$. Indica también si esta realización debe ser rechazada o no.

5. Sea una red bayesiana que consta de las variables X_1, X_2, X_3 y X_4, donde los estados posibles de cada variable son $\{0, 1\}$ para las variables X_1 y X_4, y $\{0, 1, 2\}$ para las variables X_2 y X_3; los arcos de la red son $X_1 \longrightarrow X_3$, $X_2 \longrightarrow X_3$ y $X_3 \longrightarrow X_4$. Dado el siguiente conjunto de 10 ejemplos de esas variables:

X_1	X_2	X_3	X_4
1	0	0	0
0	1	0	1
1	2	1	0
1	2	0	0
1	1	2	0
0	0	1	1
0	2	1	1
1	1	0	0
0	1	2	0
1	1	2	0

   Calcular las distribuciones de probabilidad necesarias para esa red bayesiana empleando tanto el estimador de máxima verosimilitud como el estimador bayesiano de Laplace.

6. Describe con cierto detalle las características de los métodos de aprendizaje de la estructura de una red bayesiana basados en funciones de evaluación y técnicas de búsqueda. Pon ejemplos de los distintos elementos que configuran este tipo de métodos.

7. Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, C\}$, C tiene 3 casos y el resto son binarias.

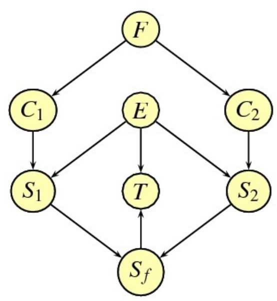
   - dibuja una red para cada una de las siguientes topologías: NB, TAN, BAN y augmented naive bayesian k-dependiente con k=2,
   - Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red
   - Ordena los modelos de menor a mayor complejidad.

8. Explica porqué RRBB aprendidas mediante métodos de ChowLiu, Friedman, o PC (basado en independencias) aplicado sin restricciones son malos clasificadores, esto es, obtienen generalmente peores tasas de clasificación que clasificadores específicos. Describe, al menos, 3 criterios diferentes utilizados durante el aprendizaje del modelo que hayan dado lugar a mejores clasificadores.

MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Curso 2017/2018, 16 de marzo de 2018

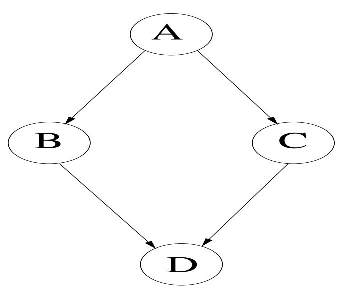
1.- En una facultad el 15 % de las personas son profesores, el 75 % son alumnos y el resto son empleados de administración y servicios. Las encuestas previas muestran que, en las últimas elecciones a Rector, votaron al candidato que al final ganó, el 55 % de los profesores, el 60 % de los alumnos y el 35 % de los empleados. Si se selecciona al azar una persona de la facultad y se sabe que no votó al candidato finalmente ganador (no votó o, si lo hizo, no lo votó a él) en las elecciones pasadas, ¿cuál es la probabilidad de que sea un profesor?

2.- Dada la siguiente red bayesiana:



Calcula razonadamente si se verifica $I(F, E|\emptyset)$, $I(F, T|E, Sf)$ e $I(F, Sf|S1, S2, T)$.

3.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0, a_1)$, $B(b_0, b_1)$, ...).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(a_1) = 0{,}7$
- $P(b_1|a_1) = 0{,}4$ y $P(b_1|a_0) = 0{,}5$
- $P(c_1|a_1) = 0{,}6$ y $P(c_1|a_0) = 0{,}9$
- $P(d_1|b_1, c_1) = 0{,}98$ y $P(d_1|b_1, c_0) = 0{,}8$
- $P(d_1|b_0, c_1) = 0{,}9$ y $P(d_1|b_0, c_0) = 0{,}0$

1. Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad, de forma que siempre aparezca primero el estado con subíndice 0 y luego el estado con subíndice 1 para cada una de las variables en cada tabla.

2. Supongamos que en todas las variables, los estados con subíndice 1 indican valor positivo o propiedad presente, y los estados con subíndice 0 valor negativo o propiedad ausente. Escribe la tabla de probabilidad que se obtendría para el potencial asociado a la distribución $P(D|B,C)$ si consideramos que viene definido con un *modelo noisy-or* donde tenemos los siguientes parámetros:

   - Cuando B está presente (o sea $B = b_1$) se causa D (o sea $D = d_1$) con probabilidad 0,8
   - Cuando C está presente (o sea $C = c_1$) se causa D (o sea $D = d_1$) con probabilidad 0,9
   - Probabilidad residual de $\lambda_0 = 0{,}0001$.

   Los números de la tabla de probabilidad pueden dejarse indicados como el producto de los números necesarios en cada caso.

4.- Supongamos la red bayesiana del ejercicio anterior (apartado primero).

1. Realiza la combinación de los potenciales asociados a las distribuciones condicionales $P(B|A)$ y $P(C|A)$. En el potencial resultante, realiza la marginalización para eliminar la variable A.

2. Usando el algoritmo de eliminación de variables, queremos obtener la distribución de probabilidad a posteriori para la variable D considerando que no tenemos ninguna observación. Suponiendo que las variables se eliminan siguiendo el orden $\{C, B, A\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo.
   Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total.

5.- Describe dos métodos alternativos para estimar la probabilidad condicionada de una variable Y dado su conjunto de variables padres X_1, X_2,...,X_m en una red bayesiana, a partir de una base de datos (completa) que contiene instancias de todas las variables presentes en la red. Aplica esos métodos para el caso de la variable Y cuyos estados posibles son $\{0, 1, 2\}$, que tiene como padres a las variables X_1 y X_2, cuyos estados posibles son $\{0, 1\}$ y $\{0, 1, 2\}$ respectivamente, y disponemos del siguiente conjunto de datos:

X_1	X_2	Y
1	0	0
0	1	0
1	2	1
1	2	0
1	1	2
0	0	1
0	2	1
1	1	0
0	1	2
1	1	2

6.- Responded a las siguientes preguntas, justificando en su caso *brevemente* las respuestas:

(a) ¿Están obligados los algoritmos de aprendizaje de redes bayesianas basados en evaluación y búsqueda utilizar técnicas de búsqueda heurística?

(b) ¿Por qué es útil que las métricas empleadas en los algoritmos de aprendizaje de redes bayesianas tengan la propiedad de "descomponibilidad"?

(c) ¿Es cierto que los algoritmos de aprendizaje de redes bayesianas basados en evaluación y búsqueda emplean diferentes métricas y diferentes técnicas de búsqueda, pero siempre realizan la búsqueda en el espacio de grafos dirigidos acíclicos?

(d) Enumera algunas métricas empleadas por los algoritmos de aprendizaje de redes bayesianas que estén basadas en ideas bayesianas y comenta brevemente la idea en la que se basan.

(e) ¿Por qué es importante para los algoritmos de aprendizaje de redes bayesianas basados en detección de independencias que los tests de independencia realizados sean del menor orden posible?

7.- Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5, C\}$, donde X_1, X_3 tienen 3 casos y el resto son binarias.

- Dibuja una red para cada una de las siguientes topologías: NB, TAN, BAN y augmented naive bayesian k-dependiente con k=2.
- Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red.
- Ordena los modelos de menor a mayor complejidad.

8.- Sean el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5, C\}$ y un conjunto de datos D_{Tr} y D_{Tst}. Suponiendo que se comprueba que $Dep(X_4, X_5|C)$ es alta, y que $P(X_1|C) = P(X_1)$, qué opinas de las tasas de éxito de clasificación del Naive Bayes respecto a las tasas de otros clasificadores tipo 1. *selective*, 2. *semi-naive* o 3. *augmented*, con los mismos conjuntos D_{Tr} y D_{Tst}. Justifica la respuesta, eligiendo 2 de los 3 tipos propuestos.

MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Curso 2018/2019, 5 de abril de 2019

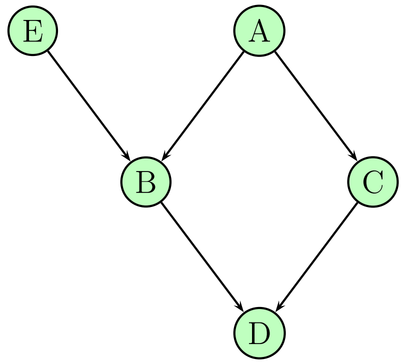
1.- Se envía un valor $Xin\{0,1\}$ por un canal de comunicación C_1. Este canal tiene una probabilidad de 0.9 de que se reciba el valor correcto y de 0.1 de que se reciba un valor erróneo (esta probabilidad no depende del valor emitido). A continuación el valor recibido en este canal, Y, se envía por otro canal de comunicación C_2 siendo Z el valor recibido en este segundo canal. Se supone que $P(Z=0|Y=0)=0{,}8$ y $P(Z=1|Y=1)=0{,}95$. Suponiendo que X y Z son condicionalmente independientes dado Y, calcula la probabilidad de que se haya recibido al final $Z=1$ si se ha emitido por el primer canal $X=1$.

2.- Quiero hacer una red bayesiana para hacer un pronóstico de la probabilidad de infarto en el próximo año (variable X). Para ello cuento con las siguientes variables:

- Obesidad (O)
- Ejercicio físico (F)
- Dieta (D)
- Fumador (S)
- Consumo Alcohol (A)
- Edad (E)
- Sexo (G)
- Estado Cardiovascular (R)
- Tensión Arterial (T)
- Electrocardiograma (C)

Dibujar un grafo dirigido acíclico que represente correctamente las independencias condicionadas existentes entre las variables del problema, teniendo en cuenta que X es independiente del resto de las variables dado R (R es la única variable que influye en X). Hacer explícitas hipótesis de independencia que se hayan usado y que resulten dudosas en el enunciado del problema.

3.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0,a_1)$, $B(b_0,b_1)$, …).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(a_1) = 0{,}5$
- $P(e_1) = 0{,}4$
- $P(b_1|e_0, a_0) = 0{,}0$, $P(b_1|e_0, a_1) = 0{,}6$, $P(b_1|e_1, a_0) = 0{,}5$, $P(b_1|e_1, a_1) = 0{,}8$
- $P(c_1|a_0) = 0{,}9$ y $P(c_1|a_1) = 0{,}6$
- $P(d_1|b_0, c_0) = 0{,}0$, $P(d_1|b_0, c_1) = 0{,}9$, $P(d_1|b_1, c_0) = 0{,}8$, $P(d_1|b_1, c_1) = 0{,}98$

1. Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad.

2. Suponiendo que usamos el algoritmo de inferencia de eliminación de variables, realiza la combinación de aquellos potenciales necesarios para poder eliminar la variable A. En el potencial resultante, realiza la marginalización para eliminar la variable A.

3. Supongamos que queremos obtener la distribución de probabilidad a posteriori para la variable D considerando que no tenemos ninguna observación en las demás variables. Suponiendo que las variables se eliminan siguiendo el orden $\{A, B, C, E\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo, solo las operaciones realizadas con potenciales, especificando las variables de cada potencial obtenido.
   Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total.

4.- La fiebre (F) puede estar causada por tener un resfriado (R), por tener gripe (G) o bien por tener la malaria (M). Supongamos que R, G y M causan fiebre (F) de forma independiente, y que la probabilidad condicionada de fiebre dadas las causas puede expresarse con un *modelo noisy-or*. Un resfriado causa fiebre con probabilidad 0,4; la gripe causa fiebre con probabilidad 0,8 y la malaria causa fiebre con probabilidad 0,9. Suponemos además que la fiebre puede estar causada por otras causas con probabilidad 0,005.

Todas las variables tienen dos estados posibles; supongamos que los estados con subíndice 1 indican valor positivo o propiedad presente, y los estados con subíndice 0 valor negativo o propiedad ausente. Por ejemplo, los estados de F son f_0 (no tener fiebre) y f_1 (sí tener fiebre).

Obten la tabla para la la distribución de probabilidad condicional de tener fiebre dadas las posibles causas, generada por el anterior *modelo noisy-or*. No es necesario calcular el resultado de la expresión usada para obtener cada probabilidad.

5.- Describe dos métodos alternativos para estimar la probabilidad condicionada de una variable Y dado su conjunto de variables padres X_1, X_2, ..., X_m en una red bayesiana, a partir de una base de datos (completa) que contiene instancias de todas las variables presentes en la red. Aplica esos métodos para el caso de la variable Y cuyos estados posibles son $\{0, 1, 2\}$, que tiene como padres a las variables X_1 y X_2, cuyos estados posibles son $\{0, 1\}$ y $\{0, 1, 2\}$ respectivamente, y disponemos del siguiente conjunto de datos:

X_1	X_2	Y
0	0	0
0	1	0
1	2	1
1	2	0
1	1	2
0	0	1
0	2	1
1	1	0
0	1	2
1	1	2

6.- Describe las características de las dos grandes clases de métodos para el aprendizaje de la estructura de una red bayesiana a partir de datos.

7.- Explica por qué RRBB aprendidas mediante los métodos de Chow-Liu, Friedman, o PC aplicados sin restricciones son malos clasificadores, esto es, obtienen generalmente peores tasas de clasificación que clasificadores específicos. Describe al menos 3 criterios diferentes utilizados durante el aprendizaje del modelo que hayan dado lugar a mejores clasificadores.

8.- Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, C\}$ todas con dos casos, excepto C con 3,

- Dibuja una red para cada una de las siguientes topologías: BAN, augmented naive bayesian k-dependiente con k=2, NB y TAN.
- Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red.
- Ordena los modelos de menor a mayor complejidad.

MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Curso 2019/2020, 30 de abril de 2020

1.- Describir un grafo dirigido acíclico que represente las dependencias/independencias del siguiente problema que relaciona la posibilidad de que una persona tenga una infección vírica en dos instantes, relacionándolo con la exposición al virus, las pruebas realizadas, la fortaleza de su sistema inmunológico y los síntomas. Las variables son:

- $X1$ la persona tiene la infección vírica en el instante 1
- $X2$ la persona tiene la infección vírica en el instante 2
- $C1$ la persona ha estado en contacto con personas infectadas en un tiempo anterior al instante 1
- $C2$ la persona ha estado en contacto con personas infectadas en un tiempo anterior al instante 2
- $T1$ la persona tiene tos en el instante 1
- $T2$ la persona tiene tos en el instante 2
- $F1$ la persona tiene fiebre en el instante 1
- $F2$ la persona tiene fiebre en el instante 2
- $PCR1$ la persona ha dado positivo en el instante 1 a una prueba PCR
- $PCR2$ la persona ha dado positivo en el instante 2 a una prueba PCR
- $A1$ la persona ha dado positivo a una prueba de anticuerpos en el instante 1
- $A2$ la persona ha dado positivo a una prueba de anticuerpos en el instante 2
- S fortaleza del sistema inmunitario

Se supone lo siguiente: si la persona está infectada en el instante 1, entonces puede seguir infectada en el instante 2, aunque no haya estado en contacto con una persona infectada entre el tiempo 1 y 2 (no se ha curado todavía). El test PCR sólo es sensible al hecho de que la persona tenga la infección en ese momento. El test de anticuerpos mide si la persona ha tenido la enfermedad en este momento o en el pasado.

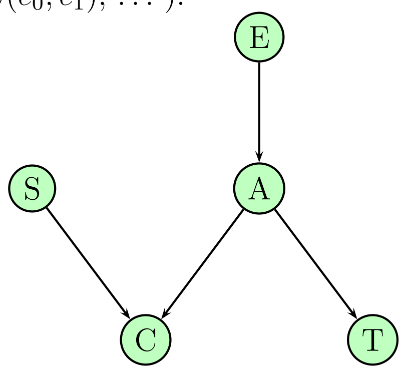
Los enfermos pueden ser asintomáticos, pero la aparición de un síntoma hace al otro más probable.

Los síntomas en el instante 2, sólo dependen de si la persona está infectada en ese instante: conocida $X2$, los síntomas son independientes del resto de las variables.

Podéis mandar una foto con el grafo, o indicar la lista de enlaces. Por ejemplo $S \rightarrow X1, ...$

Explicar si habéis realizado hipótesis adicionales.

2.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0, a_1)$, $E(e_0, e_1)$, $\ldots$).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(s_1) = 0{,}1$
- $P(e_1) = 0{,}2$
- $P(a_1|e_0) = 0{,}015$ y $P(a_1|e_1) = 0{,}0001$
- $P(c_1|s_0, a_0) = 0{,}01$, $P(c_1|s_0, a_1) = 0{,}1$, $P(c_1|s_1, a_0) = 0{,}1$, $P(c_1|s_1, a_1) = 0{,}2$
- $P(t_1|a_0) = 0{,}01$ y $P(t_1|a_1) = 0{,}3$

1. Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad.

2. Suponiendo que usamos el algoritmo de inferencia de eliminación de variables, realiza la combinación de aquellos potenciales necesarios para poder eliminar la variable E. En el potencial resultante, realiza la marginalización para eliminar la variable E.

3. Supongamos que queremos obtener la distribución de probabilidad a posteriori para la variable C considerando que no tenemos ninguna observación en las demás variables. Suponiendo que las variables se eliminan siguiendo el orden $\{T, E, S, A\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo, solo las operaciones realizadas con potenciales, especificando las variables de cada potencial obtenido.

   Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total.

3.-

(a) Describe *brevemente* dos métodos alternativos para estimar las probabilidades en una red bayesiana a partir de un conjunto de datos (completo).

(b) Sea una red bayesiana que consta de las variables X_1, X_2, X_3 y X_4, donde los estados posibles de cada variable son $\{0,1\}$ para las variables X_1 y X_4, y $\{0,1,2\}$ para las variables X_2 y X_3; los arcos de la red son $X_1 \longrightarrow X_3$, $X_2 \longrightarrow X_3$ y $X_3 \longrightarrow X_4$. Dado el siguiente conjunto de 10 ejemplos de esas variables:

X_1	X_2	X_3	X_4
1	0	0	0
0	1	0	1
1	2	1	0
1	2	0	0
1	1	2	0
0	0	1	1
0	2	1	1
1	1	0	0
0	1	2	0
1	1	2	0

Calcular las distribuciones de probabilidad necesarias para esa red bayesiana empleando los estimadores descritos previamente. Indicad los valores de probabilidad como fracciones, no como números reales.

4.- Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5, C\}$, donde X_1, X_3 tienen 3 casos y el resto son binarias.

- Dibuja una red para cada una de las siguientes topologías: NB, TAN, BAN y augmented naive bayesian k-dependiente con k=2,
- Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red
- Ordena los modelos de menor a mayor complejidad.

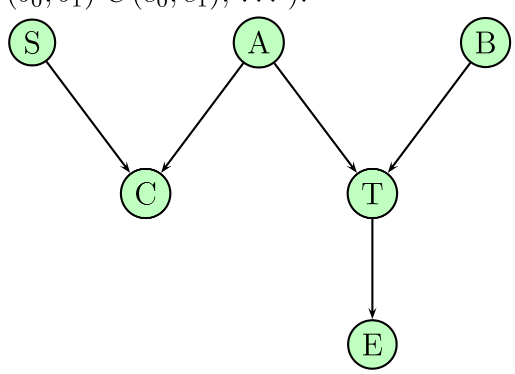
MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Convocatoria ordinaria, Curso 2020/2021
19 de abril de 2021

1.- Describir un grafo dirigido acíclico para el siguiente problema que represente correctamente sus relaciones de independencia:

Se quieren analizar dos posibles enfermedades: gripe, G, y tifus, T. Ambas producen dolor de cabeza (D) y fiebre alta (F) y dolor muscular (M). Los síntomas son más severos en el tifus que en la gripe. El tifus produce presión alterial baja (B) y la gripe no tiene efecto en la tensión. La gripe produce congestión nasal (N) y dolor faríngeo (O). La gripe se puede diagnosticar con una prueba de laboratorio (L) que es más sensible en hombres que en mujeres (X). El tifus se puede desarrollar por haber ingerido comida en mal estado (H) y la gripe depende de la estación del año (E) y de la edad (A) y de si estamos en un episodio de gripes (R).

Describe como se descompone la distribución de probabilidad conjunta como un producto de distribuciones condicionadas de acuerdo con el grafo construido.

2.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0, a_1)$, $B(b_0, b_1)$ $C(c_0, c_1)$, ...).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(s_1) = 0{,}3$
- $P(a_1) = 0{,}2$
- $P(b_1) = 0{,}1$
- $P(c_1|s_0, a_0) = 0{,}01$, $P(c_1|s_0, a_1) = 0{,}1$, $P(c_1|s_1, a_0) = 0{,}1$, $P(c_1|s_1, a_1) = 0{,}2$
- $P(t_1|a_0, b_0) = 0{,}9$, $P(t_1|a_0, b_1) = 0{,}6$, $P(t_1|a_1, b_0) = 0{,}5$ y $P(t_1|a_1, b_1) = 0{,}3$
- $P(e_1|t_0) = 0{,}4$ y $P(e_1|t_1) = 0{,}8$

1. Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad.

2. Suponiendo que usamos el algoritmo de inferencia de eliminación de variables, realiza la combinación de aquellos potenciales necesarios para poder eliminar la variable E. En el potencial resultante, realiza la marginalización para eliminar la variable E.

3. Supongamos que queremos obtener la distribución de probabilidad a posteriori para la variable C considerando que no tenemos ninguna observación en las demás variables. Suponiendo que las variables se eliminan siguiendo el orden $\{E, T, B, A, S\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo, solo las operaciones de combinación y marginalización realizadas con potenciales, especificando las variables de los potenciales que se combinan o marginalizan y las variables del potencial obtenido como resultado.

   Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total en el proceso completo para obtener la distribución a posteriori de la variable C.

3.- Responded a las siguientes preguntas, justificando *brevemente* las respuestas:

(a) ¿Están obligados los algoritmos de aprendizaje de redes bayesianas basados en evaluación y búsqueda a utilizar técnicas de búsqueda heurística?

(b) ¿Por qué es útil que las métricas empleadas en los algoritmos de aprendizaje de redes bayesianas tengan la propiedad de "descomponibilidad"?

(c) ¿Es cierto que los algoritmos de aprendizaje de redes bayesianas basados en evaluación y búsqueda emplean diferentes métricas y diferentes técnicas de búsqueda, pero siempre realizan la búsqueda en el espacio de grafos dirigidos acíclicos?

(d) Enumera algunas métricas empleadas por los algoritmos de aprendizaje de redes bayesianas que estén basadas en ideas bayesianas y comenta brevemente la idea en la que se basan.

(e) ¿Por qué es importante para los algoritmos de aprendizaje de redes bayesianas basados en detección de independencias que los tests de independencia realizados sean del menor orden posible?

(f) ¿Cuáles son las diferencias principales entre el estimador bayesiano y el de máxima verosimilitud para la estimación de probabilidades a partir de datos?

(g) ¿Es cierto que cuando disponemos de pocos datos, las diferencias entre el estimador bayesiano y el de máxima verosimilitud no son importantes?

(h) Describe (indica la fórmula) alguna versión sencilla de la estimación bayesiana a partir de datos de la probabilidad condicional $p(X = x|Y = y)$.

4.- Dado el conjunto de variables $\mathbf{X} = \{X_0, X_1, X_2, X_3, X_4, X_5, \}$, donde X_0 es la variable clase con 3 casos mientras que el resto de características son binarias.

(a) Dibuja una red bayesiana para cada una de las siguientes topologías: NB, TAN y BAN.

(b) Indica la complejidad de cada una de las redes bayesianas, medida como el número de parámetros a almacenar en ella.

(c) A igualdad de tasas de éxito en los 3 modelos que has ilustrado en el apartado anterior, indica con qué modelo te quedarías. Justifica la respuesta.

MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Convocatoria ordinaria, Curso 2021/2022
6 de abril de 2022

1.- Describir un grafo dirigido acíclico para el siguiente problema que represente correctamente sus relaciones de dependencia e independencia:

- Se quiere realizar un estudio sobre la prevalencia de dos tipos de virus: $V1$ y $V2$.

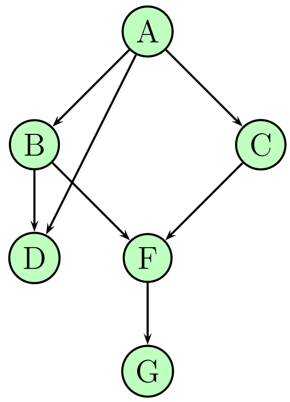
- Ambas variantes son susceptibles de producir inflamación en las articulaciones (IA) y enrojecimiento en la piel (R). La primera variante produce también trombos en las piernas (T) y algunas veces dolor al respirar (D). Por su parte, la segunda variante puede disminuir la capacidad respitatoria (CR).

- La ingesta de vegetales (V) sin limpiar puede producir la infección con ambos tipos de virus; la infeccción con la segunda variante también puede contraerse en caso de haber tenido alguna hospitalización (H).

- Las dos variantes pueden detectarse mediante un simple test de sangre (TS), pero sus resultados son más fiables en personas de más de 30 años (30+).

  Describe cómo se descompone la distribución de probabilidad conjunta como un producto de distribuciones condicionadas de acuerdo con el grafo construido.

2.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0, a_1)$, $B(b_0, b_1)$ $C(c_0, c_1)$, $\ldots$).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(a_1) = 0.5$
- $P(b_1|a_0) = 0.8$, $P(b_1|a_1) = 0.4$
- $P(c_1|a_0) = 0.1$, $P(c_1|a_1) = 0.4$
- $P(d_1|a_0, b_0) = 0.9$, $P(d_1|a_0, b_1) = 0.6$, $P(d_1|a_1, b_0) = 0.5$ y $P(d_1|a_1, b_1) = 0.3$
- $P(f_1|b_0, c_0) = 0.1$, $P(f_1|b_0, c_1) = 0.9$, $P(f_1|b_1, c_0) = 0.8$ y $P(f_1|b_1, c_1) = 0.95$
- $P(g_1|f_0) = 0.1$, $P(g_1|f_1) = 1.0$

1. Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad.
2. Suponiendo que usamos el algoritmo de inferencia de eliminación de variables, calcula la tabla de probabilidad que se obtiene como resultado de combinar aquellos potenciales necesarios para poder eliminar la variable F. Usando el potencial resultante, calcula la tabla de probabilidad que se obtiene con la operación de marginalización al eliminar la variable F.

3. Supongamos que queremos obtener la distribución de probabilidad a posteriori para la variable B considerando que no tenemos ninguna observación en las demás variables. Suponiendo que las variables se eliminan siguiendo el orden $\{G, F, D, C, A\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo, solo las operaciones de combinación y marginalización realizadas con potenciales, especificando las variables de los potenciales que se combinan o marginalizan y las variables del potencial obtenido como resultado.

   Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total en el proceso completo para obtener la distribución a posteriori de la variable B.

3.- Supongamos que tenemos dos variables X (vacunados, Si o No) e Y (han pasado Covid, con valores Si o No) y en una población hemos contado los casos de estas variables con las siguientes frecuencias:

	X=Si	X = No
Y = Si	10	20
Y = No	30	10

Estimar la distribución de probabilidad condicionada de Y dado X por los siguientes métodos:

1. Máxima verosimilitud
2. Corrección de Laplace
3. Método bayesiano con un tamaño muestral equivalente $s = 8$

4.- Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5, C\}$, donde C tienen 3 casos y el resto son binarias.

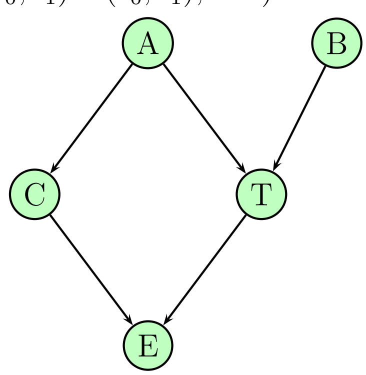
- Dibuja una red para cada una de las siguientes topologías: NB, TAN, BAN y augmented naive bayesian k-dependiente con k=2.
- Indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red.
- Ordena los modelos de menor a mayor complejidad.

MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Convocatoria ordinaria, Curso 2022/2023
24 de marzo de 2023

1.- Describir un grafo dirigido acíclico para el siguiente problema que represente correctamente sus relaciones de dependencia e independencia:

- Se considera el problema de transmitir palabras de longitud 5, formadas por el alfabeto $A = \{a, b\}$ sobre un canal de transmisión.
- Las palabras se transmiten símbolo a símbolo.
- Hay que tener en cuenta que la línea es susceptible de tener ruido, de forma que no siempre se recibe el símbolo emitido y la probabilidad de error depende únicamente del símbolo emitido.
- Se sabe que las palabras emitidas no son aleatorias, de forma que el valor de un símbolo depende del símbolo emitido con anterioridad.
- Se trata de describir una red bayesiana que represente este problema, para la emisión de una palabra completa.
- Describe también cómo se descompone la distribución de probabilidad conjunta como un producto de distribuciones condicionadas de acuerdo con el grafo construido.

2.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0, a_1)$, $B(b_0, b_1)$ $C(c_0, c_1)$, ...).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(a_1) = 0.2$
- $P(b_1) = 0.1$
- $P(c_1|a_0) = 0.1$, $P(c_1|a_1) = 0.2$
- $P(t_1|a_0, b_0) = 0.9$, $P(t_1|a_0, b_1) = 0.6$, $P(t_1|a_1, b_0) = 0.5$ y $P(t_1|a_1, b_1) = 0.3$
- $P(e_1|t_0, c_0) = 0.4$, $P(e_1|t_1, c_0) = 0.8$, $P(e_1|t_0, c_1) = 0.3$ y $P(e_1|t_1, c_1) = 0.6$

1. Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad.

2. Suponiendo que usamos el algoritmo de inferencia de eliminación de variables, realiza la combinación de aquellos potenciales necesarios para poder eliminar la variable A. En el potencial resultante, realiza la marginalización para eliminar la variable A.

3. Supongamos que queremos obtener la distribución de probabilidad a posteriori para la variable C considerando que no tenemos ninguna observación en las demás variables. Suponiendo que las variables se eliminan siguiendo el orden $\{E, T, B, A\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo, solo las operaciones de combinación y marginalización realizadas con potenciales, especificando las variables de los potenciales que se combinan o marginalizan y las variables del potencial obtenido como resultado.

   Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total en el proceso completo para obtener la distribución a posteriori de la variable C.

3.- Responde a las siguientes cuestiones:

1. Describe brevemente los dos métodos más importantes de aprendizaje de la estructura de una red bayesiana.

2. A la hora de estimar una probabilidad condicionada, ¿qué diferencia hay entre la corrección de Laplace y un método bayesiano con un tamaño muestral equivalente S?

3. ¿Qué métodos conoces para estimar probabilidades en una red bayesiana cuando hay datos perdidos? Describe brevemente sus principios.

4.- Dado el conjunto de variables $\mathbf{X} = \{X_0, X_1, X_2, X_3, X_4, C, \}$, donde C es la variable clase con 3 casos mientras que el resto de características son binarias.

- Dibuja una red bayesiana para cada una de las siguientes topologías: NB, TAN y BAN.

- Indica la complejidad de cada una de las redes bayesianas, medida como el número de parámetros a almacenar en ella.

- Ilustra el proceso de exploración de un algoritmo Semi-Naive como FSSJ (Forward Sequential Selection and Joining de Pazzani) con las variables dadas, en la que debe de figurar la que podría ser una configuración final para la red bayesiana aprendida.

MODELOS GRÁFICOS PROBABILÍSTICOS
Máster en Ciencia de Datos e Ingeniería de Computadores
Convocatoria ordinaria, Curso 2023/2024
9 de febrero de 2024

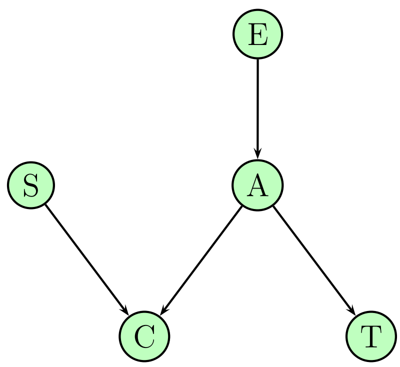
1.- La probabilidad de que una vaca sufra mastitis un determinado día depende de varios factores: si ya padecía esta dolencia el día anterior, del número de días habitual para la enfermedad y del número de días que lleva recibiendo tratamiento. La enfermedad se diagnostica mediante la observación del estado general de la vaca y de un test sobre la leche. El test puede no ser fiable si la vaca ha estado sometida a tratamiento durante más de tres días. Indica claramente:

a) las variables que intervienen en el problema, eligiendo una letra que identifique a cada una de ellas

b) describe la red Bayesiana que representa el problema usando las variables identificadas

c) indica qué distribuciones de probabilidad serían necesarias para especificar de forma completa el problema.

2.- Supongamos la siguiente red bayesiana donde todas las variables tienen dos casos ($A(a_0, a_1)$, $E(e_0, e_1)$, ...).



Supongamos que tenemos los siguientes valores de probabilidad:

- $P(s_1) = 0.1$
- $P(e_1) = 0.2$
- $P(a_1|e_0) = 0.015$ y $P(a_1|e_1) = 0.0001$

- $P(c_1|s_0, a_0) = 0.01$; $P(c_1|s_0, a_1) = 0.1$; $P(c_1|s_1, a_0) = 0.1$; $P(c_1|s_1, a_1) = 0.2$
- $P(t_1|a_0) = 0.01$ y $P(t_1|a_1) = 0.3$

a) Escribe las tablas de probabilidad para los potenciales asociados a todas las distribuciones de probabilidad.

b) Realiza la combinación de los potenciales $P(C|S, A)$ y $P(A|E)$. Suponiendo que usamos el algoritmo de inferencia de eliminación de variables, indica si tras la operación anterior, ya sería posible eliminar la variable A. En caso contrario, indica qué haría falta por hacer para poder eliminar A.

c) Supongamos que queremos obtener la distribución de probabilidad a posteriori para la variable C considerando que no tenemos ninguna observación en las demás variables. Suponiendo que las variables se eliminan siguiendo el orden $\{T, A, E, S\}$, detalla paso a paso los cálculos (combinación y marginalización de potenciales) que se realizan. No hace falta mostrar el contenido (los números) de los potenciales que se van obteniendo, solo las operaciones realizadas con potenciales, especificando las variables de cada potencial obtenido.
Indica el número de multiplicaciones y de sumas de números que es necesario realizar en cada operación con potenciales.

Indica el número de multiplicaciones y de sumas de números que es necesario realizar en total en el proceso completo para obtener la distribución a posteriori de la variable C.

3.- Explica brevemente los dos enfoques básicos para aprender la estructura de una red bayesiana. Pon un ejemplo de un algoritmo para cada uno de los enfoques describiendo de forma breve cómo funciona.

4.- Dado el conjunto de variables $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5, C\}$, donde X_2, X_4 tienen 3 casos y el resto son binarias.

- dibuja una red para cada una de las siguientes topologías: NB, TAN, BAN y augmented naive bayesian k-dependiente con k=3,
- indica la complejidad de cada una de las redes, medida como el número de parámetros a almacenar en la red y ordénalas de menor a mayor complejidad.
- indica qué modelo(s) de los anteriores no podrían ser obtenidos mediante métodos de clasificación tipo *semi-NB*, justifícalo.