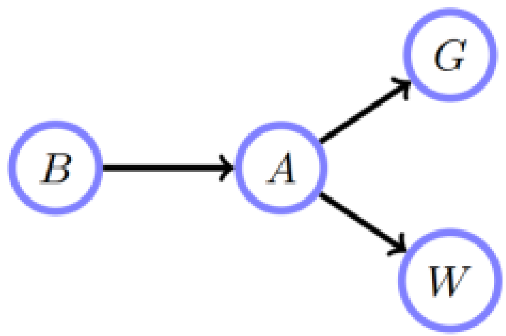
CS-E4820 Machine Learning: Advanced Probabilistic Methods (spring 2021)

Pekka Marttinen, Santosh Hiremath, Tianyu Cui, Yogesh Kumar, Zheyang Shen, Alexander Aushev, Khaoula El Mekkaoui, Shaoxiong Ji, Alexander Nikitin, Sebastiaan De Peuter, Joakim Järvinen.

Exercise 2, due on Tuesday February 2 at 23:00.

Problem 1: Computing conditional probability

Consider the Bayesian network in figure below which represents Mr Holmes' burglary worries: (B)urglar, (A)larm, (W)atson, Mrs (G)ibbon. All variables are binary with states $\{\mathsf{tr}, \mathsf{fa}\}$.



The probabilities are

$$
\begin{aligned}
p(B = \mathsf{tr}) &= 0.01 & & \\
p(A = \mathsf{tr}|B = \mathsf{tr}) &= 0.99 & \quad p(A = \mathsf{tr}|B = \mathsf{fa}) &= 0.05 \\
p(W = \mathsf{tr}|A = \mathsf{tr}) &= 0.90 & \quad p(W = \mathsf{tr}|A = \mathsf{fa}) &= 0.5 \\
p(G = \mathsf{tr}|A = \mathsf{tr}) &= 0.7 & \quad p(G = \mathsf{tr}|A = \mathsf{fa}) &= 0.2
\end{aligned}
$$

Compute the conditional probabilities

(a) $p(B = \mathsf{tr}|W = \mathsf{tr})$

(b) $p(B = \mathsf{tr}|W = \mathsf{tr}, G = \mathsf{fa})$

Write your solutions in LateX or attach a picture in the answer cell provided below. You can add a picture using the command `!(imagename_in_the_folder.jpg)`. Latex in here works similarly as you would write it normally! You can use some of the definitions from the exercise description as a reference. The list of valid Latex commands in Jypyter notebook can be found here: http://www.onemathematicalcat.org/MathJaxDocumentation/TeXSyntax.htm

Solution

Using the structure of the joint probability

$$p(B, A, G, W) = p(B)p(A|B)p(W|A)p(G|A),$$

and the definitions of conditional probability and marginalization we get

$$p(B = tr|W = tr) = \frac{p(B = tr, W = tr)}{p(W = tr)} \tag{1}$$

$$= \frac{\sum_{a,g} p(B = tr, A = a, G = g, W = tr)}{\sum_{a,b,g} p(B = b, A = a, G = g, W = tr)} \tag{2}$$

$$= \frac{p(B = tr) \sum_a p(A = a|B = tr)p(W = tr|A = a) \sum_g p(G = g|A = a)}{\substack{p(B=tr) \sum_a p(A=a|B=tr)p(W=tr|A=a) \sum_g p(G=g|A=a) + \\ p(B=fa) \sum_a p(A=a|B=fa)p(W=tr|A=a) \sum_g p(G=g|A=a)}} \tag{3}$$

$$= \frac{0.01 \times (0.99 \times 0.9 + 0.01 \times 0.5)}{\substack{0.01 \times (0.99 \times 0.9 + 0.01 \times 0.5) + \\ 0.99 \times (0.05 \times 0.9 + 0.95 \times 0.5)}} = 0.017 \tag{4}$$

and

$$p(B = tr|W = tr, G = fa) = \frac{\sum_a p(B = tr, A = a, W = tr, G = fa)}{\sum_{a,b} p(B = b, A = a, W = tr, G = fa)} \tag{5}$$

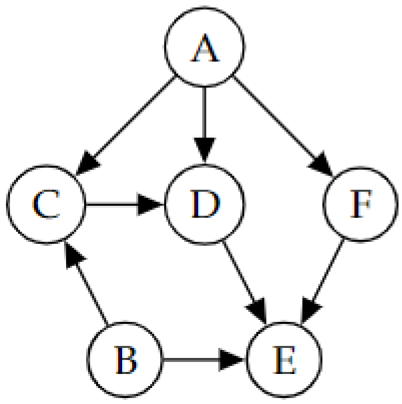
$$= \frac{p(B = tr) \sum_a p(A = a|B = tr)p(W = tr|A = a)p(G = fa|A = a)}{\substack{p(B=tr) \sum_a p(A=a|B=tr)p(W=tr|A=a)p(G=fa|A=a) + \\ p(B=fa) \sum_a p(A=a|B=fa)p(W=tr|A=a)p(G=fa|A=a)}} \tag{6}$$

$$= \frac{0.01 \times (0.99 \times 0.9 \times 0.3 + 0.01 \times 0.5 \times 0.3)}{\substack{0.01 \times (0.99 \times 0.9 \times 0.3 + 0.01 \times 0.5 \times 0.3) + \\ 0.99 \times (0.95 \times 0.9 \times 0.3 + 0.95 \times 0.5 \times 0.8)}} \tag{7}$$

$$= 0.0069 \tag{8}$$

Problem 2: Conditional independence from Bayesian network

Based on the Bayesian network in figure below, which of the following conditional independence statements follow? For each statement, give a ''true/false'' answer; for the false statements, also mention a path between the two nodes that is not blocked. (see Barber: Bayesian Reasoning and Machine Learning,ch. 3.3.4)



$$(a)\ \ A \perp\!\!\!\perp B \mid C \qquad (c)\ \ C \perp\!\!\!\perp E \mid B, D \qquad (e)\ \ B \perp\!\!\!\perp F \mid A, C \tag{9}$$

$$(b)\ \ A \perp\!\!\!\perp B \mid \emptyset \qquad (d)\ \ C \perp\!\!\!\perp D \mid A, B \qquad (f)\ \ A \perp\!\!\!\perp E \mid D, F \tag{10}$$

Furthermore, find a Bayesian network that is *Markov equivalent* to the network in the above figure. (see Barber: Bayesian Reasoning and Machine Learning,ch. 3.3.6)

Write your solution in LateX or attach a picture of the solution in the cell below.

Solution

Task		*Blocked*	*Path*
a	$A \perp\!\!\!\perp B \mid C$	*No*	Path $A \to C \to B$ is not blocked
b	$A \perp\!\!\!\perp B \mid \emptyset$	*Yes*	All paths are blocked
c	$C \perp\!\!\!\perp E \mid B, D$	*No*	Path $C \to A \to F \to E$ is not blocked
d	$C \perp\!\!\!\perp D \mid A, B$	*No*	Path $C \to D$ is not blocked
e	$B \perp\!\!\!\perp F \mid A, C$	*Yes*	All paths are blocked
f	$A \perp\!\!\!\perp E \mid D, F$	*No*	Path $A \to C \to B \to E$ is not blocked
			(as D is an observed descendant of C)

The only Markov equivalent network is found by reversing the direction of the edge $A \to F$.

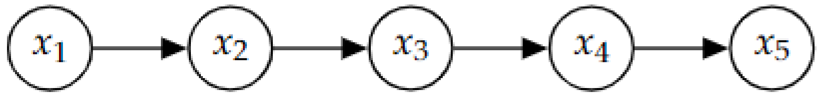
Problem 3: Burden of specification

Consider a distribution of five binary variables x_i.

(a) What is the number of parameters needed to define the distribution $p(x_1, x_2, x_3, x_4, x_5)$ if no assumptions are made, i.e. p is an arbitrary distribution.

(b) How about if the Bayesian network in figure below is assumed, i.e. p factorizes as implied by the graph.

(c) And how about if, additionally to (b), we assume that the conditional distributions are shared, i.e. $p(x_{i+1} \mid x_i) = p(x_i \mid x_{i-1})$, $i = 2, 3, 4$?



Write your solution in LateX or attach a picture of the solution in the cell below.

Solution

(a) In the generic case, the joint distribution $p(x_1, x_2, x_3, x_4, x_5)$ can be represented by 31 scalar parameters: there are $2^5 = 32$ possible states the five binary variables can be in, and the normalization constraint $\sum p(x_1, x_2, x_3, x_4, x_5) = 1$ can be used to reduce the number by one.

As an alternative derivation, consider (without loss of generality) the factorization: $p(x_1, x_2, x_3, x_4, x_5) = p(x_1 \mid x_2, x_3, x_4, x_5)p(x_2 \mid x_3, x_4, x_5)p(x_3 \mid x_4, x_5)p(x_4 \mid x_5)p(x_5)$. To define e.g. $p(x_2 \mid x_3, x_4, x_5)$, 8 parameters are needed: $p(x_2 = 1 \mid x_3, x_4, x_5)$ for every possible value of x_3, x_4, x_5. Therefore $16 + 8 + 4 + 2 + 1 = 31$ parameters suffice to define the full joint distribution.

(b) Given the structure implied by the network in the above figure, the joint distribution can be factorized as $p(x_1, x_2, x_3, x_4, x_5) = p(x_1)\, p(x_2 \mid x_1) \cdots p(x_5 \mid x_4)$. It takes a single scalar value to define $p(x_1)$, and two to define any of the factors $p(x_{i+1} \mid x_i)$, giving a total of $1+2+2+2+2 = 9$ parameters.

(c) As before, but there is only a single distribution for $p(x_{i+1} \mid x_i)$, so the total number of parameters is $1 + 2 = 3$.

Problem 4: Medical diagnosis

Let's have the following notation:

Notation	Explanation
$A = 1$	A person has brain cancer
$B = 1$	A person has a high blood calcium level
$C = 1$	A person has a brain tumor
$D = 1$	A person has seizures that cause unconsciousness
$E = 1$	A person has severe headaches

An expert has told us the following information about the relationships between variables:

> Probability of severe headaches $P(E = 1)$ depends only on the fact whether a person has a brain tumor (C) or not. On the other hand, if one knows the blood calcium level (B) and whether the person has a tumor or not (C), one can specify the probability of unconsciousness seizures $P(D = 1)$. In this case, the probability of D doesn't depend on the presence of the headaches (E) or (directly) on the fact whether the person has brain cancer or not (A). The probability of a brain tumor (C) depends directly only on the fact, whether the person has brain cancer or not (A).

Construct a DAG which represents (exactly) the causal statements specified by the expert. Furthemore, write down the conditional independencies corresponding the description of the expert and verify that this conforms with those implied by the DAG. Finally, write down the joint distribution $P(A, B, C, D, E)$ factorized according to the DAG.

Write your solution in LateX or attach a picture of the solution in the cell below.

Solution

From the description given by the expert, we get the following DAG:

The conditional independencies corresponding to the expert's description are:

$E \perp\!\!\!\perp A \mid C$
$E \perp\!\!\!\perp B \mid C$
$E \perp\!\!\!\perp D \mid C$
$D \perp\!\!\!\perp E \mid \{B, C\}$ (already implicit in the above statement)
$D \perp\!\!\!\perp A \mid \{B, C\}$
$C \perp\!\!\!\perp B \mid A$

Note that C is a direct cause of D and E, which is why $C \not\perp\!\!\!\perp D \mid A$ and $C \not\perp\!\!\!\perp E \mid A$. Also note that the expert did not say anything about a possible causal dependence between A and B. However, the above conditional independence statements would remain the same even if we added an edge between A and B (in either direction).

The joint distribution factorized according to the DAG is:

$P(A, B, C, D, E) = P(A)\ P(B)\ P(C \mid A)\ P(D \mid B, C)\ P(E \mid C)$.