csc384probquiz

https://www.notion.so/frankcyh/csc384probquiz-5f184048f86340699fb7

Question 1

Question

Bayes' Burrito Bowls is a food truck that serves burrito bowls (obviously). You can choose:

- a type of rice, which can be; white, brown
- a type of bean, which can be; black, pinto, lima, or kidney
- any of the following toppings; guacamole, corn, salsa, sour cream, onion, lettuce, or cheese

We model each type of burrito as a triple (R, B, T), where R and B are the types of rice and bean, and T is any subset of the toppings. The type of burrito a customer picks depends on his/her preferences. We can model such preferences using a probability distribution over (R, B, T).

Answer

(a) How many different types of burrito bowls are there?

To determine the number of different types of burrito bowls, we multiply the number of choices for each component:

Rice (R): 2 types (white, brown)

Beans (B): 4 types (black, pinto, lima, kidney)

Toppings (T): Since there are 7 toppings and each can either be chosen or not, there are 2^7 combinations

In total, there are

$$2 \times 4 \times 2^7 = 1024$$

possible combinations of toppings.

(b) Eshan builds a burrito bowl as follows: He first chooses the type of rice and bean independently of each other. Based on both these choices, he chooses the toppings (the toppings are not chosen independently of each other). Provide an appropriate factorization of $P_{\text{Eshan}}(R, B, T)$ and the number of values that must be known to compute it for any R, B, T.

Eshan builds a burrito bowl by choosing the type of rice and bean independently, and then, based on both these choices, he chooses the toppings. The toppings are not chosen independently of each other, but their choice depends on the chosen rice and beans.

The appropriate factorization for Eshan's choice model is:

$$P_{\text{Eshan}}(R, B, T) = P(R) \times P(B) \times P(T \mid R, B)$$

To compute this for any R, B, T, we need to know:

- The probabilities of choosing each type of rice, $P(R)$, which requires 2 values (since the probability of one can be determined knowing the other).

- The probabilities of choosing each type of bean, $P(B)$, which requires 4 values (similarly, knowing three allows us to determine the fourth).
- The probabilities of choosing each combination of toppings given each combination of rice and bean types, $P(T \mid R, B)$. Since there are 2 choices for rice, 4 for beans, and 2^7combinations of toppings, this requires $2 \times 4 \times 2^7 = 1024$ values.
- So, in total, we need $2 + 4 + 1024 = 1030$ values to fully specify $P_{\text{Eshan}}(R, B, T)$.

(c) Zafeer builds a book as follows: He first chooses type of rice. Based on this choice, he chooses a type of bean. He chooses the toppings independently of the other choices (toppings are not chosen independently of each other). Provide an appropriate factorization of $P_{\text{Zafeer}}(R, B, T)$ and the number of values that must be known to compute it for any R, B, T.

Zafeer builds a burrito bowl by first choosing the type of rice, then based on this choice, he chooses a type of bean. He chooses the toppings independently of the rice and bean choices (but toppings are not chosen independently of each other).

The appropriate factorization for Zafeer's choice model is:

$$P_{\text{Zafeer}}(R, B, T) = P(R) \times P(B \mid R) \times P(T)$$

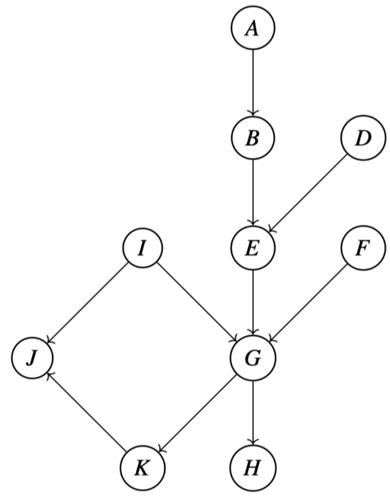
To compute this for any R, B, T, we need to know:

- The probabilities of choosing each type of rice, $P(R)$, which requires 2 values.
- The probabilities of choosing each type of bean given the rice choice, $P(B \mid R)$, which requires $2 \times 4 = 8$ values since there are 2 rice choices and 4 possible bean choices for each rice type.
- The probabilities of choosing each combination of toppings, $P(T)$, which requires $2^7 = 128$ values since the toppings are chosen independently of the rice and bean choices.
- So, in total, we need $2 + 8 + 128 = 138$ values to fully specify $P_{\text{Zafeer}}(R, B, T)$.

Question 2

Question

Consider the Bayesian network shown below



Answer

(a) Which variables are independent of I ?

$$A, B, D, E, F$$

A, B, D, E, F all belongs to the third situation in d-separation

(b) Which variables are conditionally independent of A given K ?

The descendants of $A(B, E, G, H)$ are not independent of A.

D is not independent of A because K is the descendant of E.

F, I, J are not independent of A because K is the descendant of G.

A, B, D, E, F all belongs to the third situation in d-separation

Therefore, there are no variables that are conditionally independent of A given K.

(c) Which variables are conditionally independent of B given E ?

When E is known, it blocks any influence B might have on its descendants because E is a common child of B and D. Therefore, given E, B is conditionally independent of F, G, H, J, K, I (since E intercepts the only path B has to influence these variables).

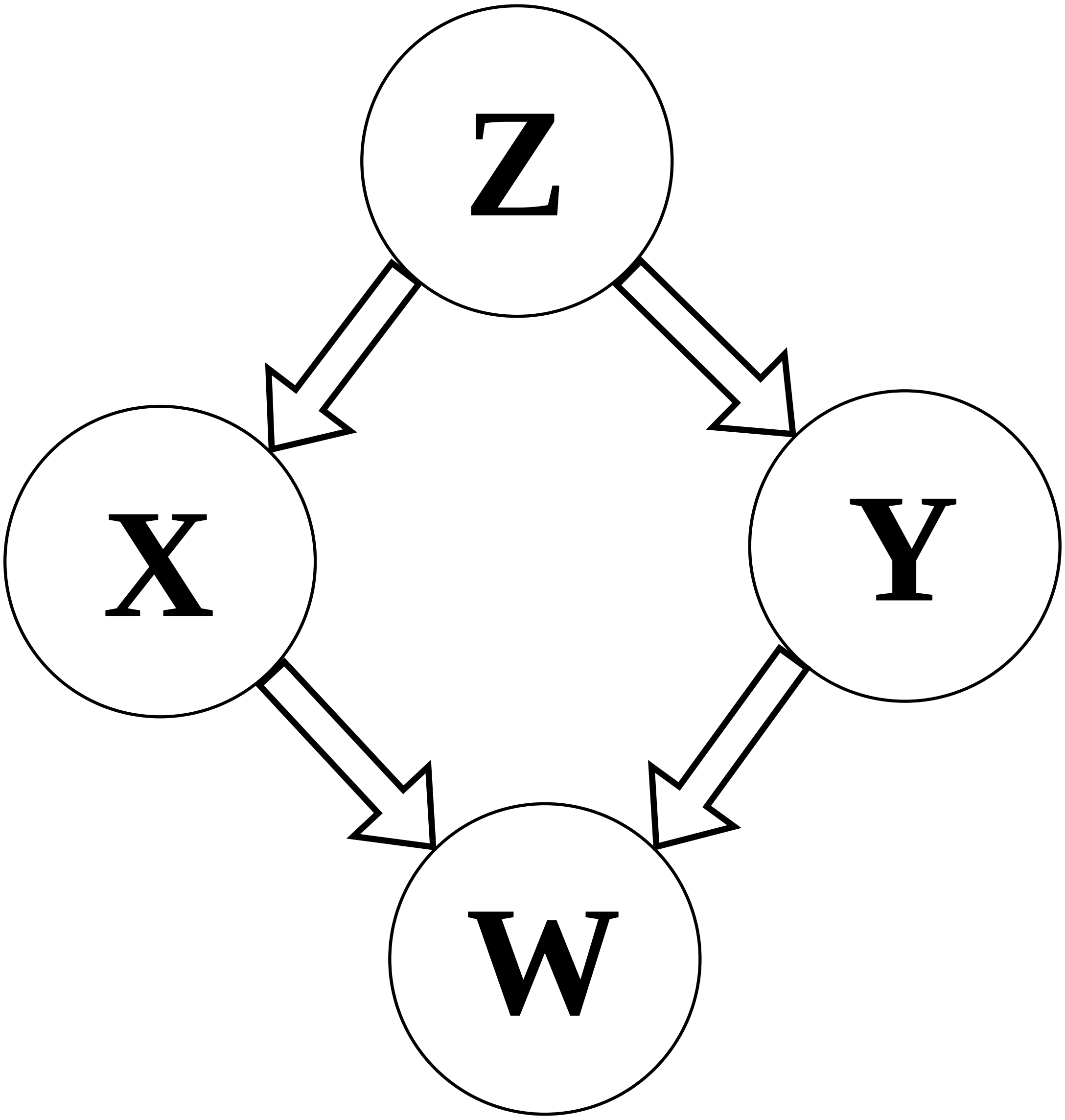
D is not conditionally independent of B because it violate the third rule of d-separation

Therefore, the variables that are conditionally independent of B given E are F, G, H, I, J, K.

Question 3

Draw a Bayesian network over the variables W, X, Y, Z, such that:

1. X and Y are independent given Z
2. Z and W are independent given X and Y

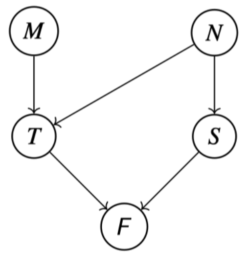


Question 4

Question

You and your friend are on the Amazing Race (see Wikipedia article for details) and the next leg is in a few hours. Once the leg begins, you will need to make your way to the airport. You want to use this time to decide how you are going to get there, and how many bags you need to pack.

$P(M = \text{train})$
0.6



M	N	$P(T = \text{early} \mid M, N)$
train	0	0.9
train	1	0.85
train	2	0.60
car	0	0.75
car	1	0.75
car	2	0.75

$P(N = 0)$	$P(N = 1)$
0.4	0.5

N	$P(S = \text{fast} \mid N)$
0	0.9
1	0.8
2	0.7

T	S	$P(F = \text{yes} \mid T, S)$
early	fast	0.95
early	slow	0.7
late	fast	0.7
late	slow	0.35

Both of these decisions influence whether you will make your flight or miss it by influencing when you get to the airport and how long it takes you to get through security. This scenario is modelled using the Bayesian network above:

- $M \in \{\text{train,car}\}$ is how you will get to the airport
- $N \in \{0, 1, 2\}$ is the number of bags you pack
- $T \in \{\text{early, late}\}$ is when you get to the airport
- $S \in \{\text{fast, slow}\}$ is how long it takes to get through security
- $F \in \{\text{yes, no}\}$ is whether you make the flight or not

Answer

(a) Compute $P(F = \text{yes} \mid N, M)$ for all M and N. Show your work.

$M = \text{train}, N = 0$

$$
\begin{aligned}
&P(F = \text{yes} \mid T = \text{early}, S = \text{fast}) \times P(S = \text{fast} \mid N = 0) \times P(T = \text{early} \mid M = \text{train}, N = 0) + \\
&P(F = \text{yes} \mid T = \text{early}, S = \text{slow}) \times P(S = \text{slow} \mid N = 0) \times P(T = \text{early} \mid M = \text{train}, N = 0) + \\
&P(F = \text{yes} \mid T = \text{late}, S = \text{fast}) \times P(S = \text{fast} \mid N = 0) \times P(T = \text{late} \mid M = \text{train}, N = 0) + \\
&P(F = \text{yes} \mid T = \text{late}, S = \text{slow}) \times P \mid S = \text{slow} \mid N = 0) \times P(T = \text{late} \mid M = \text{train}, N = 0) \\
&= 0.95 \times 0.9 \times 0.9 + 0.7 \times 0.1 \times 0.9 + 0.7 \times 0.9 \times 0.1 + 0.35 \times 0.1 \times 0.1 \\
&= 0.8990
\end{aligned}
$$

$M = \text{train}, N = 1$

$P(F=\text{yes} \mid T=\text{early}, S=\text{fast}) \times P(S=\text{fast} \mid N=1) \times P(T=\text{early} \mid M=\text{train}, N=1)+$
$P(F=\text{yes} \mid T=\text{early}, S=\text{slow}) \times P(S=\text{slow} \mid N=1) \times P(T=\text{early} \mid M=\text{train}, N=1)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{fast}) \times P(S=\text{fast} \mid N=1) \times P(T=\text{late} \mid M=\text{train}, N=1)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{slow}) \times P(S=\text{slow} \mid N=1) \times P(T=\text{late} \mid M=\text{train}, N=1)$
$=0.95 \times 0.8 \times 0.85+0.7 \times 0.2 \times 0.85+0.7 \times 0.8 \times 0.15+0.35 \times 0.2 \times 0.15$
$=0.8595$

$M=\text{train}, N=2$

$P(F=\text{yes} \mid T=\text{early}, S=\text{fast}) \times P(S=\text{fast} \mid N=2) \times P(T=\text{early} \mid M=\text{train}, N=2)+$
$P(F=\text{yes} \mid T=\text{early}, S=\text{slow}) \times P(S=\text{slow} \mid N=2) \times P(T=\text{early} \mid M=\text{train}, N=2)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{fast}) \times P(S=\text{fast} \mid N=2) \times P(T=\text{late} \mid M=\text{train}, N=2)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{slow}) \times P(S=\text{slow} \mid N=2) \times P(T=\text{late} \mid M=\text{train}, N=2)$
$=0.95 \times 0.7 \times 0.6+0.7 \times 0.3 \times 0.6+0.7 \times 0.7 \times 0.4+0.35 \times 0.3 \times 0.4$
$=0.7630$

$M=\text{car}, N=0$

$P(F=\text{yes} \mid T=\text{early}, S=\text{fast}) \times P(S=\text{fast} I N=0) \times P(T=\text{early} \mid M=\text{car}, N=0)+$
$P(F=\text{yes} \mid T=\text{early}, S=\text{slow}) \times P(S=\text{slow} I N=0) \times P(T=\text{early} \mid M=\text{car}, N=0)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{fast}) \times P(S=\text{fast} I N=0) \times P(T=\text{late} \mid M=\text{car}, N=0)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{slow}) \times P \mid S=\text{slow} \mid N=0) \times P(T=\text{late} \mid M=\text{car}, N=0)$
$=0.95 \times 0.9 \times 0.75+0.7 \times 0.1 \times 0.75+0.7 \times 0.9 \times 0.25+0.35 \times 0.1 \times 0.25$
$=0.86$

$M=\text{car}, N=1$

$P(F=\text{yes} \mid T=\text{early}, S=\text{fast}) \times P(S=\text{fast} \mid N=1) \times P(T=\text{early} \mid M=\text{car}, N=1)+$
$P(F=\text{yes} \mid T=\text{early}, S=\text{slow}) \times P(S=\text{slow} \mid N=1) \times P(T=\text{early} \mid M=\text{cal}, N=1)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{fast}) \times P(S=\text{fast} \mid N=1) \times P(T=\text{late} \mid M=\text{car}, N=1)+$
$P(F=\text{yes} \mid T=\text{late}, S=\text{slow}) \times P \mid S=\text{slow} \mid N=1) \times P(T=\text{late} \mid M=\text{car}, N=1)$
$=0.95 \times 0.8 \times 0.75+0.7 \times 0.2 \times 0.75+0.7 \times 0.8 \times 0.25+0.35 \times 0.2 \times 0.25$
$=0.8325$

$M=\text{car}, N=2$

$$
\begin{aligned}
&P(F=\text{yes} \mid T=\text{early}, S=\text{fast}) \times P(S=\text{fast} I N=2) \times P(T=\text{early} \mid M=\text{car}, N=2)+ \\
&P(F=\text{yes} \mid T=\text{early}, S=\text{slow}) \times P(S=\text{slow} \mid N=2) \times P(T=\text{early} \mid M=\text{cal}, N=2)+ \\
&P(F=\text{yes} \mid T=\text{late}, S=\text{fast}) \times P(S=\text{fast} \mid N=2) \times P(T=\text{late} \mid M=\text{car}, N=2)+ \\
&P(F=\text{yes} \mid T=\text{late}, S=\text{slow}) \times P(S=\text{slow} \mid N=2) \times P(T=\text{late} \mid M=\text{car}, N=2) \\
&=0.95 \times 0.7 \times 0.75+0.7 \times 0.3 \times 0.75+0.7 \times 0.7 \times 0.25+0.35 \times 0.3 \times 0.25 \\
&=0.805
\end{aligned}
$$

(b) How should you get to the airport, and how many bags should you bring to maximize your chances of making the flight?

From the calculation above, the optimal strategy is to choose the $M=\text{train}, N=0$, which has a possibility of 0.8990.