CS 440 Homework Assignment 3 Part 1
Probabilistic Reasoning

Devvrat Patel and Shubham Mittal

November 20, 2019

Problem 1. : Consider the following Bayesian network, where variables A through E are all Boolean valued.

a) What is the probability that all five of these Boolean variables are simultaneously true?

For this question we will use the fact that

$$P(A, B, C, D, E) = P(A) * P(B) * P(C) * P(D|A, B) * P(E|B, C).$$

$$\begin{aligned} P(A = T, B = T, C = T, D = T, E = T) &= P(A = T) * P(B = T) * P(C = T) * P(D = T) * P(E = T) \\ &= 0.2 * 0.5 * 0.8 * 0.1 * 0.3 \\ &= 0.0024 \end{aligned}$$

b) What is the probability that all five of these Boolean variables are simultaneously false?

$$\begin{aligned} P(A = F, B = F, C = F, D = F, E = F) &= P(A = F) * P(B = F) * P(C = F) * P(D = F) * P(E = F) \\ &= 0.8 * 0.5 * 0.2 * 0.1 * 0.8 \\ &= 0.0064 \end{aligned}$$

c) What is the probability that A is false given that the four other variables are all known to be true?

$$P(\neg A|B,C,D,E)$$

$$\alpha * P(\neg A, B, C, D, E)$$

$$Here,\ \alpha = \frac{1}{P(A,B,C,D,E) + P(\neg A,B,C,D,E)}$$

$$\alpha = \frac{1}{(0.2*0.5*0.8*0.1*0.3) + (0.8*0.5*0.8*0.6*0.3)}$$

$$\alpha = \frac{1}{0.0024 + 0.0576}$$

$$\alpha = \frac{50}{3}$$

$$P(\neg A|B,C,D,E)$$

$$= \alpha * P(\neg A, B, C, D, E)$$

$$= \frac{50}{3} * 0.0576$$

$$= 0.96$$

Problem 2. a) Calculate P(Burglary—JohnsCalls = true, MaryCalls = true) and show in detail the calculations that take place. Use your book to confirm that your answer is correct.

For this question we will be using the fact that,

$$
\begin{aligned}
P(B|J,M) &= \frac{P(B,J,M)}{P(J,M)} \\
&= \alpha * P(B) \sum_E P(E) \sum_A P(A|B,E) * P(J|A) * P(M|A) \ Here,\ (\alpha = \frac{1}{P(J,M)}) \\
&= \alpha * P(B) \sum_A P(E) * ((0.9 * 0.7 * \begin{pmatrix} 0.95 & 0.29 \\ 0.94 & 0.001 \end{pmatrix} + 0.5 * 0.01 * \begin{pmatrix} 1-0.95 & 1-0.29 \\ 1-0.94 & 1-0.001 \end{pmatrix})) \\
&= \alpha * P(B) \sum_A P(E) * ((0.9 * 0.7 * \begin{pmatrix} 0.95 & 0.29 \\ 0.94 & 0.001 \end{pmatrix} + 0.5 * 0.01 * \begin{pmatrix} 0.05 & 0.71 \\ 0.06 & 0.999 \end{pmatrix}) \\
&= \alpha * P(B) \sum_A P(E) * \begin{pmatrix} 0.598525 & 0.183055 \\ 0.59223 & 0.011295 \end{pmatrix} \\
&= \alpha * P(B) * (0.002 * \begin{pmatrix} 0.598525 \\ 0.183055 \end{pmatrix} + 0.998 * \begin{pmatrix} 0.59223 \\ 0.0011295 \end{pmatrix} \\
&= \alpha * \begin{pmatrix} 0.001 \\ 0.999 \end{pmatrix} * \begin{pmatrix} 0.59224259 \\ 0.001493351 \end{pmatrix} \\
&= \alpha * \begin{pmatrix} 0.00059224259 \\ 0.0014918576 \end{pmatrix} \\
&(We\ can\ calculate\ \alpha\ the\ using\ the\ same\ method.\ \alpha = \frac{1}{0.0020853609}) \\
&= \frac{1}{0.0020853609} * \begin{pmatrix} 0.00059224259 \\ 0.0014918576 \end{pmatrix} \\
&= < 0.284, 0.716 >
\end{aligned}
$$

Here, 0.284 is the probability that a burglary will happen if John and Mary calls whereas 0.716 is the probability that burglary will not happen if they both call.

b)What is the complexity of computing $P(X_1—X_n = true)$ using enumeration? What is the complexity with variable elimination?

We use enumeration for this example to compute $P(X_1—X_n = true)$. So first, we have to evaluate two binary trees for each value of X_1 and each of them have the depth n - 2. Therefore, the total work for enumeration will be $O(2^n)$.

Now, let's move to variable elimination. In variable elimination, the factors will never grow more than two variables. For example,

$$\begin{aligned}
P(X_1|X_{\text{n}} = True) &= \alpha * P(X) \;...\sum_{x_{\text{n-2}}} P(x_{\text{n-2}}|x_{\text{n-3}}) \sum_{x_{\text{n-1}}} P(x_{\text{n-1}}|x_{\text{n-2}}) P(X_{\text{n}} = True|x_{\text{n-1}}) \\
&= \alpha * P(X) \;...\sum_{x_{\text{n-2}}} P(x_{\text{n-2}}|x_{\text{n-3}}) \sum_{x_{\text{n-1}}} f X_{\text{n-1}}(x_{\text{n-1}}, x_{\text{n-2}}) f X_{\text{n}}(x_{\text{n-1}}) \\
&= \alpha * P(X) \;...\sum_{x_{\text{n-2}}} P(x_{\text{n-2}}|x_{\text{n-3}}) f \frac{x_{\text{n-2}}}{X_{\text{n-1}} * X_{\text{n}}}
\end{aligned}$$

As we see here, this is isomorphic to the problem n-1 variables and not n. Therefore the work done will be a constant independent of n and the total work is O(n).

Problem 3. Suppose you are working for a financial institution and you are asked to implement a fraud detection system.

a) Construct a Bayes Network to identify fraudulent transactions.

b) What is the prior probability (i.e., before we search for previous computer related purchases and before we verify whether it is a foreign and/or an internet purchase) that the current transaction is a fraud? What is the probability that the current transaction is a fraud once we have verified that it is a foreign transaction, but not an internet purchase and that the card holder purchased computer related accessories in the past week?

OC : card holder owns a computer or smart phone.

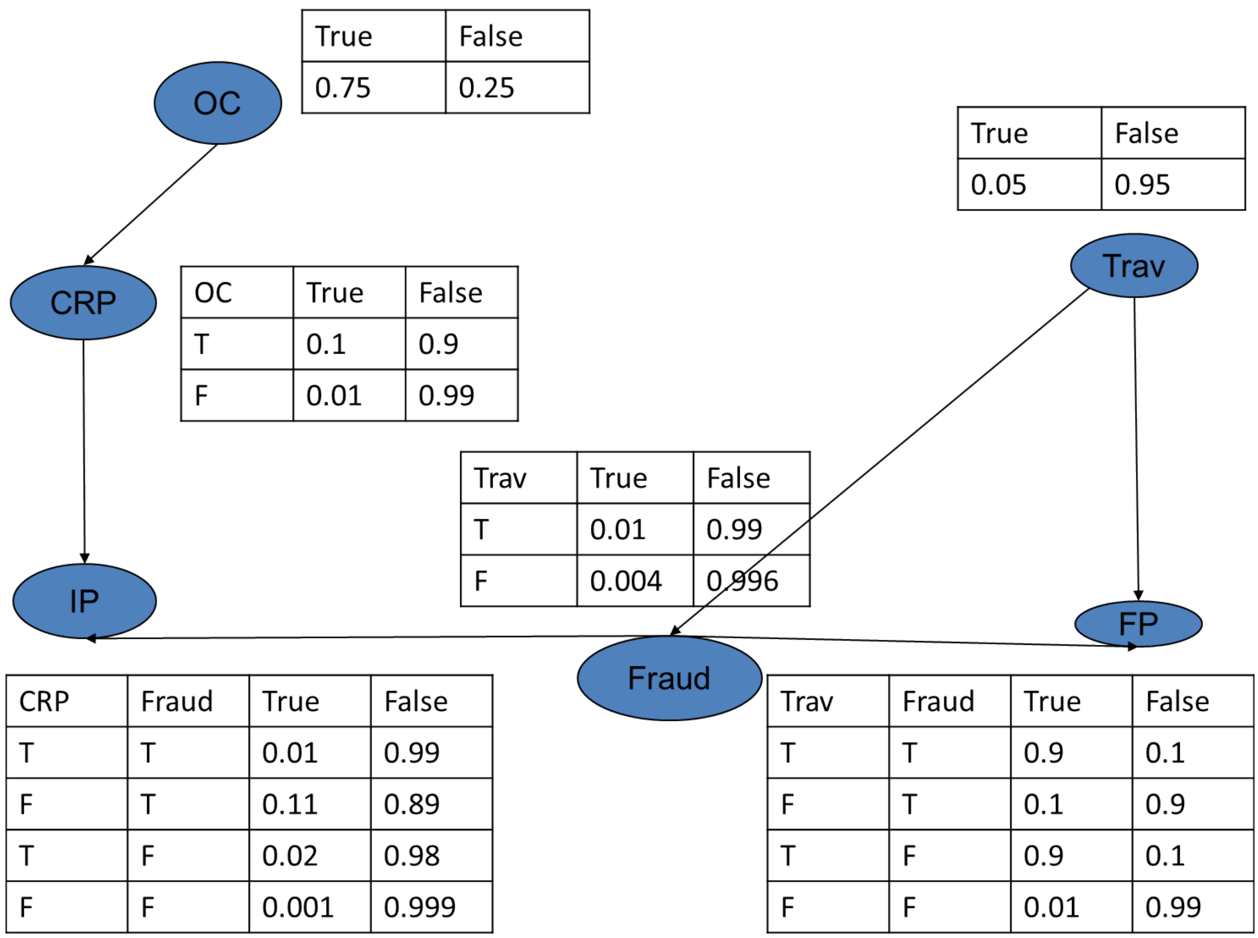
Fraud : current transaction is fraudulent.

Trav : card holder is currently travelling.

FP : current transaction is a foreign purchase.

IP : current purchase is an internet purchase.

CRP : a computer related purchase was made in the past week.



True	False
0.75	0.25

True	False
0.05	0.95

OC	True	False
T	0.1	0.9
F	0.01	0.99

Trav	True	False
T	0.01	0.99
F	0.004	0.996

CRP	Fraud	True	False
T	T	0.01	0.99
F	T	0.11	0.89
T	F	0.02	0.98
F	F	0.001	0.999

Trav	Fraud	True	False
T	T	0.9	0.1
F	T	0.1	0.9
T	F	0.9	0.1
F	F	0.01	0.99

Figure 1: Bayes Network to identify fraudulent transactions.

$$
\begin{aligned}
P(Fraud) &= P(Fraud = T|Trav = T) * P(Trav = T) + P(Fraud = T|Trav = F) * P(Trav = F) \\
&= 0.01 * 0.05 + 0.004 * 0.95 \\
&= 0.004275
\end{aligned}
$$

Let the probability that the current transaction is a fraud once we have verified that it is a foreign transaction, but not an internet purchase and that the card holder purchased computer related accessories in the past week P!

$$
\begin{aligned}
P(Fraud|FP) &= P(Fraud|Trav)P(FP|Trav, Fraud)P(Trav)+ \\
&+ P(Fraud|\neg Trav)P(FP|\neg Trav, Fraud)P(\neg Trav) \\
&= 0.01 * 0.90 * 0.05 + 0.004 * 0.10 * 0.95 \\
&= 0.00045 + 0.00038 \\
&= 0.00083
\end{aligned}
$$

$$
\begin{aligned}
P(Fraud|\neg IP, CRP, OC) &= P(\neg IP|CRP, Fraud)P(CRP)+ \\
&+ P(\neg IP|\neg CRP, Fraud)P(\neg CRP) \\
&= P(\neg IP|CRP, Fraud)* \\
&* (P(CRP|OC) * P(OC) + P(CRP|\neg OC) * P(\neg OC))+ \\
&+ P(\neg IP|\neg CRP, Fraud)* \\
&* (P(\neg CRP|OC) * P(OC) + P(\neg CRP|\neg OC) * P(\neg OC)) \\
&= 0.99 * (0.1 * 0.75 + 0.01 * 0.25) + 0.89 * (0.9 * 0.75 + 0.99 * 0.25) \\
&= 0.99 * 0.02125 + 0.89 * 0.41625 \\
&= 0.379
\end{aligned}
$$

Answer. P = 0.379 * 0.00083 = 0.00031457