

CSE 573 Problem Set 3

Due Mon 11/10/08 4:40pm

Preferably, turn in by email to both **daipeng@cs** & **weld@cs**

If any problem doesn't contain enough information for you to answer it, feel free to make any assumptions necessary to find the answer, but state the assumptions clearly. You will be graded on your choice of assumptions as well as the clarity of your written answers.

You are allowed to discuss problems with your classmates to the extent of formulating ideas. When it comes to writing up solutions, I expect each person to do their own work. I'll also request that you wait an hour between discussions of a problem and the subsequent writeup.

On the homework you submit, please include a list of all the people with whom you discussed the problem. If you consult written materials other than the course materials, please list these resources as well. (Normally, this should be unnecessary). On programming assignments, I expect you to write the code from scratch given the materials we have listed. (E.g., on PS2, please don't search for a unification routine on the Web - I'm sure that there are many out there). If you wish to write your PDDL domain by modifying an existing domain, that can be fine, but list the domain(s) you started from and include a listing.

We'll accept late problem sets under the following conditions:

- Up to 1 hour late - no penalty
- Up to one day late - 25% penalty
- Up to two day late - 50% penalty
- More than 2 days - no credit (please plan ahead!)

1. (5 points) R&N problem 13.1

$$P(a | a \wedge b) = \frac{P(a \wedge (a \wedge b))}{P(a \wedge b)} = \frac{P(a \wedge a \wedge b)}{P(a \wedge b)} = \frac{P(a \wedge b)}{P(a \wedge b)} = 1$$

2. (5 points) R&N problem 13.8

We denote a person has disease by d , and a person tested positive by p .

$$P(d | t) = \frac{P(t | d)P(d)}{P(t | d)P(d) + P(t | \neg d)P(\neg d)} = \frac{.99 \times .0001}{.99 \times .0001 + .01 \times .9999} = 9.804 \times 10^{-3}$$

3. (15 points) R&N problem 13.11

(a) We denote a fake coin by f , and the observation of a head by h

$$P(f | h) = \frac{P(h | f)P(f)}{P(h | f)P(f) + P(h | \neg f)P(\neg f)} = \frac{1/n}{(n-1)/n \times 1/2 + 1/n} = \frac{2}{n+1}$$

(b) We denote the observation of head at time t by h_t

$$P(f | h_1 \wedge \dots \wedge h_k) = \frac{P(h_1 \wedge \dots \wedge h_k | f)P(f)}{P(h_1 \wedge \dots \wedge h_k | f)P(f) + P(h_1 \wedge \dots \wedge h_k | \neg f)P(\neg f)}$$

$$= \frac{1/n}{[(n-1)/n](1/2)^k + 1/n} = \frac{2^k}{n-1+2^k}$$

(c) Here, an error occurs either in the form of a “false positive”, where we think we use a faked coin while actually it is a normal one, or a “false negative”, where the judgment and the reality is the other way around.

$$P(\text{error}) = P(h_1 \wedge \dots \wedge h_k | \neg f)P(\neg f) + P(\neg(h_1 \wedge \dots \wedge h_k) | f)P(f)$$

$$= \frac{n-1}{n} \left(\frac{1}{2}\right)^k + 0 = \frac{n-1}{n} \left(\frac{1}{2}\right)^k$$

4. (5 points) R&N problem 13.16

This is a tricky problem made famous by Monte Hall on the TV show “Let’s Make a Deal” and even professional statisticians have gotten it wrong. One common-sense argument suggests that since B is pardoned, both A and C have the same posterior probability of getting executed. In fact, the guard’s message doesn’t affect A ’s chances, so the real person to get nervous is C who now has a $2/3$ probability of impending death. We can see this formally, by denoting the guard’s message to B by mb , and the random variables corresponding to the execution of A , B or C by ea , eb , and ec . We consider the following three conditional probabilities:

$P(mb|ea)=0.5$, since the guard can tell either B or C that he will be pardoned;

$P(mb|eb=0)$, since the guard cannot tell B he is pardoned while he is actually going to be executed;

$P(mb|ec)=1$, since C will be executed, B is the only person from B or C that will be pardoned.

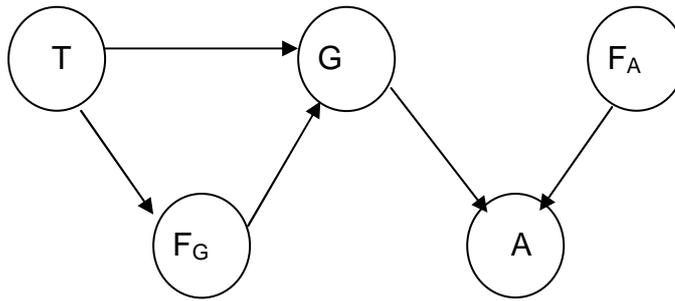
$$P(ea | mb) = \frac{P(mb | ea)P(ea)}{P(mb | ea)P(ea) + P(mb | eb)P(eb) + P(mb | ec)P(ec)}$$

$$= \frac{(1/2) \times (1/3)}{(1/2) \times (1/3) + 0 \times (1/3) + 1 \times (1/3)} = \frac{1}{3}$$

So, by knowing the guard telling B that he will be pardoned, A ’s chances of survival have not changed

5. (25 points) R&N problem 14.2

(a) The Bayesian network is shown in the graph



(b) No, it is not a polytree, since there is an undirected cycle in the network. Even if one considers a different plausible graph, it can't be a polytree because temperature affects the gauge in two different ways.

(c) The conditional probability table looks like the following:

F_G	T	$P(G=\text{normal})$	$P(G=\text{high})$
working	low	x	1-x
working	high	1-x	x
faulty	low	y	1-y
faulty	high	1-y	y

(d) The conditional probability table looks like the following:

F_A	G	$P(A=\text{on})$	$P(G=\text{off})$
working	low	0	1
working	high	1	0
faulty	low	0	1
faulty	high	0	1

(e) This part actually asks the students to do something usually done by Bayesian network algorithms. The great thing is that doing the calculation without a Bayesian network makes it easy to see the nature of the calculations that the algorithms are systematizing. It illustrates the magnitude of the achievement involved in creating complete and correct algorithms.

Abbreviating $T=High$ and $G=High$ by T and G , the probability of interest here

is $P(T \mid \neg F_G, \neg F_A)$. Because the alarm's behavior is deterministic, we can reason that if the alarm is working and sounds, G must be *High*. Because F_A and A are d-separated from T , we need only calculate $P(T \mid \neg F_G, G)$.

There are several ways to go about doing this. The "opportunistic" way is to notice that the CPT entries give us $P(T \mid \neg F_G, G)$, which suggests using the generalized Bayes' Rule to

Switch G and T with $\neg F_G$ as background:

$$P(T | \neg F_G, G) \propto P(G | T, \neg F_G)P(T | \neg F_G)$$

We then use Bayes' rule again on the last term :

$$P(T | \neg F_G, G) \propto P(G | T, \neg F_G)P(\neg F_G | T)P(T)$$

A similar relationship holds for $\neg T$:

$$P(\neg T | \neg F_G, G) \propto P(G | \neg T, \neg F_G)P(\neg F_G | \neg T)P(\neg T)$$

Normalizing, we obtain :

$$P(T | \neg F_G, G) = \frac{P(G | T, \neg F_G)P(\neg F_G | T)P(T)}{P(G | T, \neg F_G)P(\neg F_G | T)P(T) + P(G | \neg T, \neg F_G)P(\neg F_G | \neg T)P(\neg T)}$$

The “systematic” way to do it is to revert to joint entries (noticing that the sub-graph of T , G , and F_G is completely connected so no loss of efficiency is entailed). We have

$$P(T | \neg F_G, G) = \frac{P(T, \neg F_G, G)}{P(\neg F_G, G)} = \frac{P(T, \neg F_G, G)}{P(T, \neg F_G, G) + P(\neg T, \neg F_G, G)}$$

Now we use the chain rule formula (Equation 15.1 on page 439) to rewrite the joint entries as CPT entries :

$$P(T | \neg F_G, G) = \frac{P(G | T, \neg F_G)P(\neg F_G | T)P(T)}{P(G | T, \neg F_G)P(\neg F_G | T)P(T) + P(G | \neg T, \neg F_G)P(\neg F_G | \neg T)P(\neg T)}$$

which of course is the same as the expression arrived at above. Letting $P(T) = p$, $P(F_G | T) = g$, and $P(F_G | \neg T) = h$, we get :

$$P(T | \neg F_G, G) = \frac{p(1-g)(1-x)}{p(1-g)(1-x) + (1-p)(1-h)x}$$

6. (20 points) R&N problem 14.7 parts a (10 points), b (5 points) and c (5 points)
 (a) Here we want to calculate $P(b|j,m)$

$$f_M(A) = \begin{pmatrix} P(m | a) \\ P(m | \neg a) \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.01 \end{pmatrix}$$

$$f_J(A) = \begin{pmatrix} P(j | a) \\ P(j | \neg a) \end{pmatrix} = \begin{pmatrix} 0.9 \\ 0.05 \end{pmatrix}$$

The factor $f_A(A,b,E)$ is given by:

A	B	E	$P(A B,E)$
a	b	e	.95
a	b	$\neg e$.94
a	$\neg b$	e	.29
a	$\neg b$	$\neg e$.001
$\neg a$	b	e	.05
$\neg a$	b	$\neg e$.06
$\neg a$	$\neg b$	e	.71
$\neg a$	$\neg b$	$\neg e$.999

The first step is to sum out A and compute $f_{\bar{A}JM}^-(B, E)$

$$\begin{aligned}
 f_{\bar{A}JM}^-(B, E) &= f_A(a, B, E) \times f_J(a) \times f_M(a) + f_A(\neg a, B, E) \times f_J(\neg a) \times f_M(\neg a) \\
 &= \begin{pmatrix} .95 & .29 \\ .94 & .001 \end{pmatrix} \times .9 \times .7 + \begin{pmatrix} .05 & .71 \\ .06 & .999 \end{pmatrix} \times .05 \times .01 \\
 &= \begin{pmatrix} .598525 & .183055 \\ .59223 & .0011295 \end{pmatrix}
 \end{aligned}$$

Then we process E in the same way and compute $f_{\bar{E}AJM}^-(b, E)$:

$$\begin{aligned}
 f_{\bar{E}AJM}^-(b, E) &= f_E(e) \times f_{\bar{A}JM}^-(b, E) \\
 &= .002 \times \begin{pmatrix} .598525 \\ .183055 \end{pmatrix} + .998 \times \begin{pmatrix} .59223 \\ .0011295 \end{pmatrix} \\
 &= \begin{pmatrix} .59233 \\ .00113 \end{pmatrix}
 \end{aligned}$$

Finally, we can compute the final answer

$$\begin{aligned}
 P(b | j, m) &= \alpha \times f_B(b) \times f_{\bar{E}AJM}^-(b, E) \\
 &= \alpha \times \begin{pmatrix} .0001 \times .59224259 \\ .999 \times .001493 \end{pmatrix} \\
 &= \alpha \times \begin{pmatrix} .00059224259 \\ .001491857649 \end{pmatrix} = \langle .284, .716 \rangle
 \end{aligned}$$

(b) Including normalization, there are 16 multiplications, 7 additions and 2 divisions.

(c) Using enumeration algorithm, we first multiply every possibility together and sum over variables one by one. The number of multiplications is $O(2^{n-1})$, and there are $O(2^{n-1}-1)$ additions, so the time complexity is $O(2^n)$. However, if we use variable elimination, each potential function contains at most 2 variables, so it takes constant time to generate them and summing up variables. Since there are n variables, the total time is $O(n)$.

7. (6 points) R&N problem 20.4
- (a) $P(A)=.4$, $P(B)=.6$. The Bayesian statistician will take both anti- A and anti- B , since his hypothesis of his problem is a weighted combination of diseases that have probability greater than 1. The ML statistician will, on the other hand, only take anti- B , since his hypothesis is one single disease that has the highest probability, namely B .
- (b) $P(A)=.4$, $P(B_1)=.3$, $P(B_2)=.3$. In this case, the strategy of the Bayesian statistician remains unchanged, since although the hypothesis distribution has diversified, he will still consider all possibilities into consideration. For the ML statistician, he will take anti- A instead, since disease A now becomes the one that has the highest probability.

8. (30 points) Consider a Bayesian network that has three Boolean-valued nodes, A , B , and C . Assume that we have little knowledge about this network and so decide to use a Beta prior with both α and β equal to 5 for all random variables. We then observed the system, recording data about its behavior, which is stored in the following file:

(<http://www.cs.washington.edu/education/courses/573/08au/problems/ps3/data.txt>) Note that 0 denotes false and 1 denotes true.

- a) (5 points) Consider the two network structures shown on the next page. Use the data to update the prior to compute the conditional probability tables for the nodes in structure 1.

$$P(A) = \frac{43}{90} = .4778,$$

$$P(B/A) = \frac{14}{48} = .2917, \quad P(B/\neg A) = \frac{45}{52} = .8654,$$

$$P(C/A) = \frac{23}{48} = .4791, \quad P(C/\neg A) = \frac{35}{52} = .6731$$

- b) (5 points) Do likewise for structure 2.

$$P(A) = \frac{43}{90} = .4778,$$

$$P(B/A) = \frac{14}{48} = .2917, \quad P(B/\neg A) = \frac{45}{52} = .8654,$$

$$P(C/A, B) = \frac{8}{19} = .4211, \quad P(C/A, \neg B) = \frac{20}{39} = .5128,$$

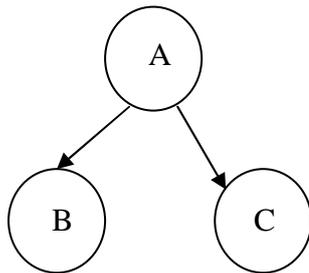
$$P(C/\neg A, B) = \frac{35}{50} = .7, \quad P(C/\neg A, \neg B) = \frac{5}{12} = .4167$$

- c) (5 points) What is the probability of observing the data given structure 1 and the learned CPT values?

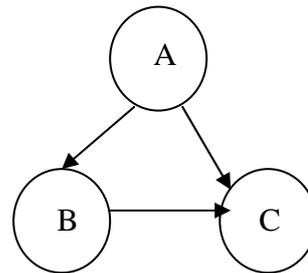
We assume that each data entry is independent given its Bayesian network structure. So

$$\begin{aligned}
P(\text{data} | \text{model}) &= \log \prod_{\text{each } d} P(d | \text{model}) \\
&= \sum_{\text{each } d} \log P(d | \text{model}) \\
&= \sum_{\text{each } d} [\log P(d_a) + \log P(d_b | d_a) + \log P(d_c | d_a)] \\
&= -137.819
\end{aligned}$$

d) (5 points) Likewise, what's the probability of observing the data for structure 2?



Structure 1



Structure 2

$$\begin{aligned}
P(\text{data} | \text{model}) &= \log \prod_{\text{each } d} P(d | \text{model}) \\
&= \sum_{\text{each } d} \log P(d | \text{model}) \\
&= \sum_{\text{each } d} [\log P(d_a) + \log P(d_b | d_a) + \log P(d_c | d_a, d_b)] \\
&= -136.014
\end{aligned}$$

e) (5 points) Which structure better models the data?

The second one, since it has higher $P(\text{data} | \text{model})$.

f) (5 points) Suppose that you are using the following utility function, U , to evaluate the two structures. Which structure do you prefer then?

$$\begin{aligned}
U &= \log(P(D|BN)) - \text{penalty}, \text{ where} \\
\text{Penalty} &= 1/2 * (\# \text{ of parameters}) * \log(\# \text{ data points}).
\end{aligned}$$

$$U_1 = -137.819 - \frac{1}{2} \times 5 \times \log 80 = -148.773$$

$$U_2 = -136.014 - \frac{1}{2} \times 7 \times \log 80 = -151.351$$

We prefer the first model now, since it has higher utility value.