

CS760 Spring 2019 Homework 2

Due Mar 7 at 11:59pm

Name: Xinyi Li

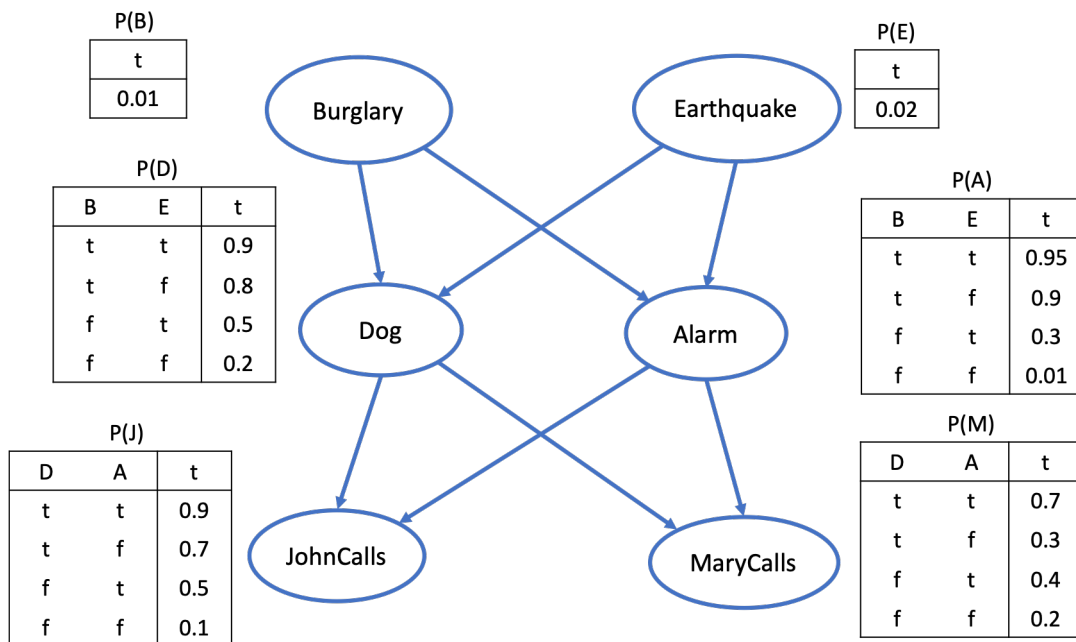
Email: xli646@wisc.edu

Written Problems

NOTE: For the following written problems, put your answer in `hw2.pdf`. You are required to provide detailed solutions including the intermediate results for each step. Otherwise, you will not get full credit. You can also add figures or tables whenever necessary. If your solutions are handwritten, make sure they are legible.

- (8 pts) Suppose you have a Bayesian network with 6 binary random variables shown as follows, where t and f stand for *true* and *false* respectively.

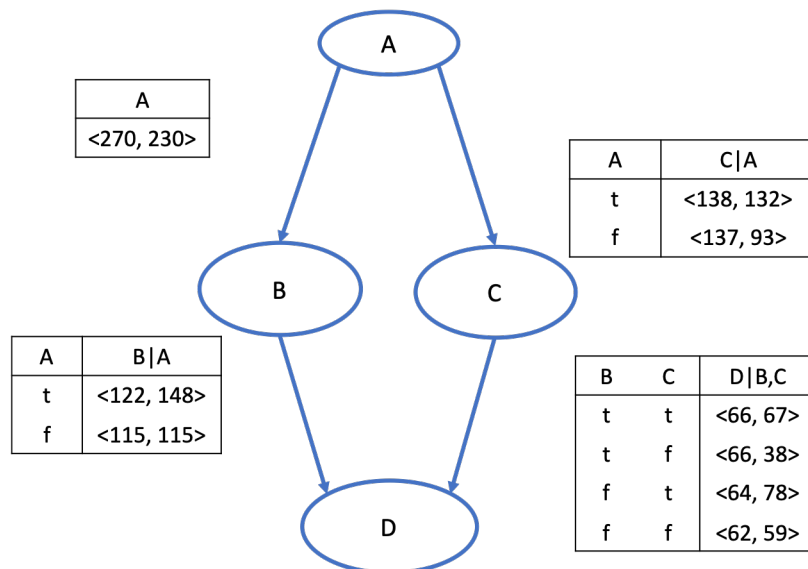
Compute the probability: $P(d|b, \neg a, j, m)$.



$$\begin{aligned}
P(d|b, \neg a, j, m) &= \frac{P(d, b, \neg a, j, m, E)}{P(D, b, \neg a, j, m, E)} \\
&= \frac{\sum_{e, \neg e} P(d, b, \neg a, j, m, E)}{\sum_{d, \neg d, e, \neg e} P(D, b, \neg a, j, m, E)} \\
&= \frac{\sum_{e, \neg e} P(b)P(E)P(d|b, E)P(\neg a|b, E)P(j|d, \neg a)P(m|d, \neg a)}{\sum_{d, \neg d, e, \neg e} P(b)P(E)P(D|b, E)P(\neg a|b, E)P(j|D, \neg a)P(m|D, \neg a)} \tag{1} \\
&= (0.01 \times 0.02 \times 0.9 \times 0.05 \times 0.7 \times 0.3 + 0.01 \times 0.98 \times 0.8 \times 0.1 \times 0.7 \times 0.3) \\
&\quad / (0.01 \times 0.02 \times 0.9 \times 0.05 \times 0.7 \times 0.3 + 0.01 \times 0.98 \times 0.8 \times 0.1 \times 0.7 \times 0.3 \\
&\quad + 0.01 \times 0.02 \times 0.1 \times 0.05 \times 0.1 \times 0.2 + 0.01 \times 0.98 \times 0.2 \times 0.1 \times 0.1 \times 0.2) \\
&= 0.9769
\end{aligned}$$

2. Given the following Bayesian network and sample counts in each table, where sample counts $\langle n_{\text{true}}, n_{\text{false}} \rangle$, there are n_{true} samples with true labels and n_{false} samples with false labels for this attribute. For example, $\langle 138, 132 \rangle$ in table $C|A$ says given the condition of $A = \text{true}$, there are 138 instances are true and 132 are false with regard to attribute C .

You need to answer the following two questions.



- (a) (2 pts) Construct the conditional probability tables (CPTs) based on the above sample count tables, using maximum likelihood estimation. You need to both show the true probability P_{true} and false probability P_{false} for each case, and organize them in the format of $\langle P_{\text{true}}, P_{\text{false}} \rangle$. For example, for the case $Y|X_1, X_2$, your answer will look like $\langle P(Y|X_1, X_2), P(\neg Y|X_1, X_2) \rangle$. Keep **at least 3 digits of precision**. (You may reuse the same structure as the above tables, just plugging in the conditional probabilities in the place of sample counts. For more information, please refer to the lecture notes `BNS-1.pdf`)

Since we are only given a set of data. Using maximum likelihood estimation to estimate CPTs is equivalently to calculate CPTs from distribution of data sets and also, we can independently calculate each CPT:

$$P(A) = \left\langle \frac{270}{500}, \frac{230}{500} \right\rangle = \langle 0.540, 0.460 \rangle$$

A
$\langle 0.540, 0.460 \rangle$

For $P(B|A)$:

A	B A
t	$\langle 0.415, 0.548 \rangle$
f	$\langle 0.500, 0.500 \rangle$

For $P(C|A)$:

A	C A
t	$\langle 0.511, 0.489 \rangle$
f	$\langle 0.596, 0.404 \rangle$

For $P(D|B, C)$:

B	C	D B,C
t	t	$\langle 0.496, 0.504 \rangle$
t	f	$\langle 0.635, 0.365 \rangle$
f	t	$\langle 0.451, 0.549 \rangle$
f	f	$\langle 0.512, 0.288 \rangle$

- (b) (10 pts) Show the result of one cycle of the EM algorithm to update the CPTs you derived in step (a), using 10 another instances with $A=\text{true}$, $B=\text{false}$, $C=?$, and $D=\text{true}$ ('?' means missing value). Keep **at least 2 digits of precision**.

E-step:

According to CPTs from part(a),

$$\begin{aligned}
P(c|a, \neg b, d) &= \frac{P(c, a, \neg b, d)}{P(c|a, \neg b, d) + P(\neg c|a, \neg b, d)} \\
&= \frac{P(a)P(c|a)P(\neg b|a)P(d|c, \neg b)}{P(a)P(c|a)P(\neg b|a)P(d|c, \neg b) + P(a)P(\neg c|a)P(\neg b|a)P(d|\neg c, \neg b)} \\
&= \frac{0.540 \times 0.511 \times 0.500 \times 0.451}{0.540 \times 0.511 \times 0.500 \times 0.451 + 0.540 \times 0.489 \times 0.500 \times 0.512} \\
&= 0.479
\end{aligned} \tag{2}$$

correspondingly:

$$P(\neg c|a, \neg b, d) = 0.521 \tag{3}$$

A	B	C	D
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t
t	f	t:0.479 f:0.521	t

M-step: Restimate the CPTs according to data distribution and here is new distribution:

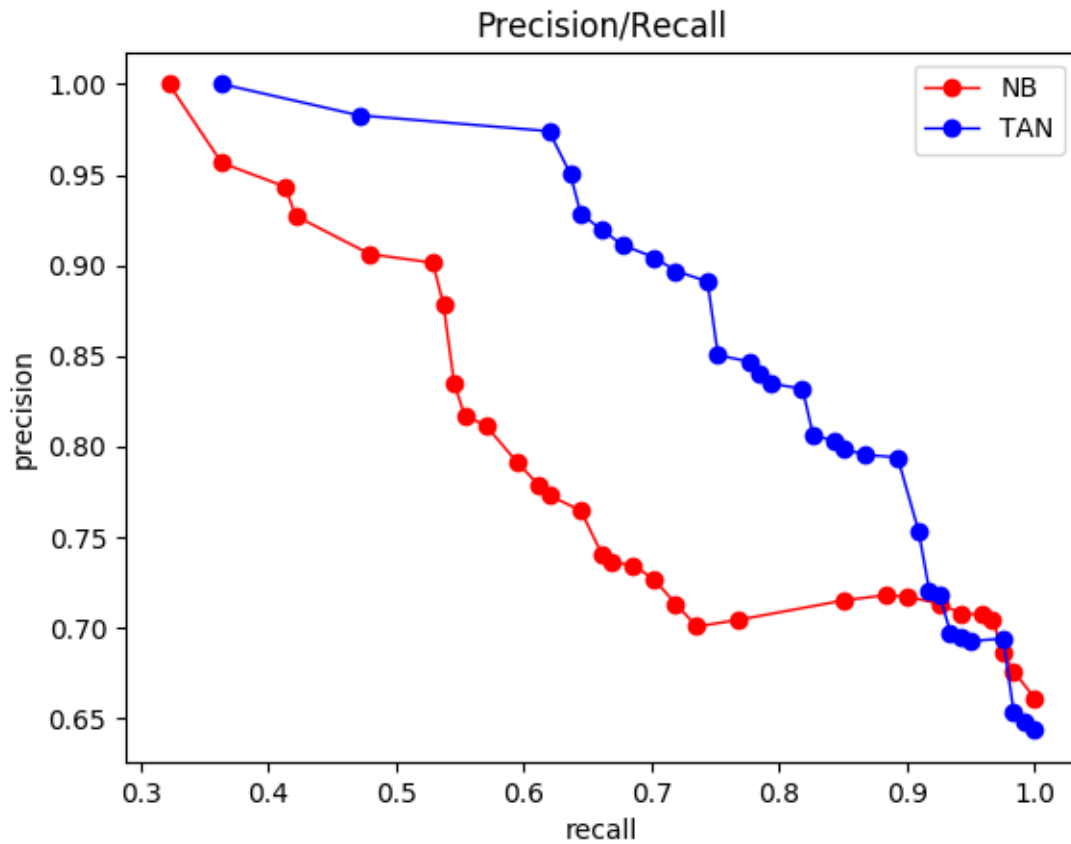
A	A	B A	A	C A	B C	D B,C
<280,230>	t	<122,158>	t	<142.79,137.21>	t t	<66,67>
	f	<115,115>	f	<137,93>	t f	<66,38>
					f t	<68.79,78>
					f f	<67.21,59>

The corresponding CPTs:

A	A	B A	A	C A	B C	D B,C
<0.549,0.451>	t	<0.436,0.564>	t	<0.510,0.490>	t t	<0.496,0.504>
	f	<0.500,0.500>	f	<0.596,404>	t f	<0.635,365>
					f t	<0.469,0.531>
					f f	<0.533,0.467>

Part 2

I think TAN seems to have more predictive power, since the area under the TAN curve is larger than the area under NB curve. In another words, TAN curve is more close to point(recall=1, precision=1).



Part 3

I shuffled the data before 10-fold cross validation:

The sample mean for NB is 0.693.

The sample mean for TAN is 0.759.

t -value is 6.654.

The corresponding t -value for a two-tailed t -test at $p = 0.05$ and degrees of freedom = 9 is 2.262.

Obtained t -value is larger than 2.262. Thus two systems seem to provide different accuracies. (TAN is better).

We can also calculate p -value corresponding to $t=6.654$, which is 0.000093, much smaller than 0.5.