CS760 Spring 2019 Homework 2

Due Mar 7 at 11:59pm

Name: Lane Schultz

Email: lsschultz@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.

1. (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.

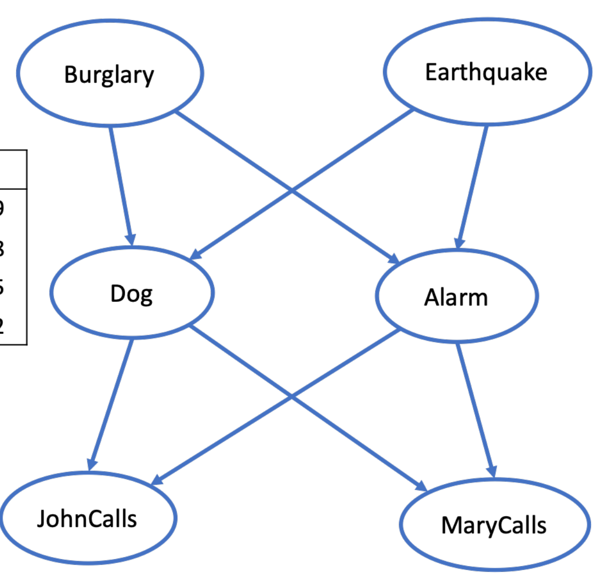
   Compute the probability: $P(d|b, \neg a, j, m)$.

P(B)

t
0.01

P(E)

t
0.02



P(D)

B	E	t
t	t	0.9
t	f	0.8
f	t	0.5
f	f	0.2

P(A)

B	E	t
t	t	0.95
t	f	0.9
f	t	0.3
f	f	0.01

P(J)

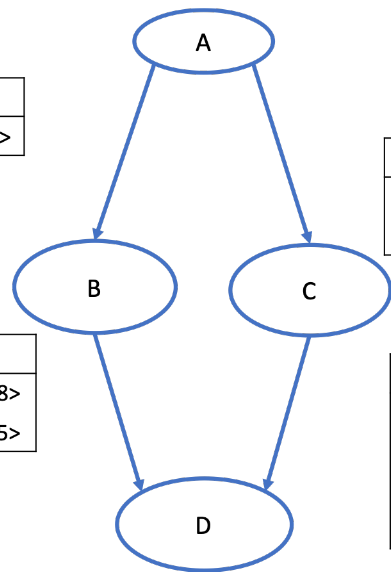
D	A	t
t	t	0.9
t	f	0.7
f	t	0.5
f	f	0.1

P(M)

D	A	t
t	t	0.7
t	f	0.3
f	t	0.4
f	f	0.2

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

You need to answer the following two questions.



A
<270, 230>

A	C|A
t	<138, 132>
f	<137, 93>

A	B|A
t	<122, 148>
f	<115, 115>

B	C	D|B,C
t	t	<66, 67>
t	f	<66, 38>
f	t	<64, 78>
f	f	<62, 59>

(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 $\mathtt{P_{true}}$ and false probability P_{false} for each case, and organize them in the format of $\mathtt{<P_{true},\ P_{false}>}$. For example, for the case $\mathtt{Y|X_1,X_2}$, your answer will look like $\mathtt{<P(Y|X_1,X_2),\ P(\neg Y|X_1,X_2)>}$. 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`)

(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=true`, `B=false`, `C=?`, and `D=true` ('?' means missing value). Keep **at least 2 digits of precision**.

For written portion refer to the written portion

Part 2 PR Curve Answer

My computed values for the PR curve are incorrect. With that being said, I can speak about what would hypothetically occur.

The TAN should have higher predictive power than the naive Bayes should. The reason for this is that TAN incorporates the influence of other features on their probability of occurring. This is why we calculate the conditional mutual information. Conversely, the naive method only looks at the probabilities of each feature with respect to the classes individually.

If it is true that TAN is superior to the naive method (which I need to assume to answer this question because of my incorrect values), the the the PR curve for TAN will be less concave upward than the naive method. In other words, TAN should have a closer number of predicted positives to that of the actual positives than the naive method.

Part 3

The t values where calculated with the following formula where m is the sample mean, μ is population mean, σ is the sample standard deviation, and n is the sample size:

$$t = (m - \mu)/(\sigma/\sqrt{n})$$

The data set was randomized at the beginning. For each cross fold, the accuracies for the naive and TAN Bayes method were calculated as the number of correct predictions over the number of samples for the test data. The TAN method accruacies were subtracted from the naive method for each cross fold. The average of the differences were considered the population mean, their standard deviation as the sample standard deviation, each delta was considered as a sample mean, and the number of subtractions was used for n. This is represented with the following code snip:

```
deltas = [i-j for i, j in zip(naiveaccuracies, tanaccuracies)]
mean = np.mean(deltas)
sig = np.std(deltas)
tvals = [(i-mean)/(sig/len(deltas)**0.5) for i in deltas]
print(tvals)
```

The values from the calculation are below. According to the p test chart provided in https://www.statisticshowto.datasciencece distribution-table/, the t value for a test with alpha being 0.05 is around 12.706. None of the absolute values of the t values calculated are above the aforementioned t value. Hence, the methods are not significantly different. This is probably because I have some errors in the coding sections.

-0.8419238590788608
1.7655012953626021
1.113645006752238
-6.056774167961794
-3.449349013520331
0.4617887181418672
1.7655012953626021
3.069213872583337
-2.866109176342632
5.038506028700972

(#1) Clark Schultz

$P(d|b, \neg a, j, m)$ is the query

From the graph structure we can first compute:

$$P(d, b, \neg a, j, m) = \prod_{i=1}^{n} P(X_i | parents(X_i))$$

$$= \sum_{e, \neg e} P(d|b,E) P(b) P(\neg a|b,E) P(j|d, \neg a) P(m|d, \neg a)$$

$$= P(b) \sum_{e, \neg e} P(d|b,E) P(\neg a|b,E) P(j|d, \neg a) P(m|d, \neg a)$$

$$\begin{array}{lllllll} & B & D & A & J & M & \\ = & 0.01 \cdot (& 0.9 & \cdot (1-0.95) & \cdot\ 0.7 & \cdot\ 0.3\ + & e \\ & & 0.8 & \cdot (1-0.9) & \cdot\ 0.7 & \cdot\ 0.3) & \neg e \end{array}$$

$$= 2.625 \cdot 10^{-4}$$

Now compute:

$$P(\neg d, b, \neg a, j, m) = \prod_{i=1}^{n} P(X_i | parents(X_i))$$

$$= \sum_{e, \neg e} P(\neg d|b,E) P(b) P(\neg a|b,E) P(j|\neg d, \neg a) P(m|\neg d, \neg a)$$

$$= P(b) \sum_{e, \neg e} P(\neg d|b,E) P(\neg a|b,E) P(j|\neg d, \neg a) P(m|\neg d, \neg a)$$

$$\begin{array}{lllllll} & B & D & A & J & M & \\ = & 0.01 \cdot (& (1-0.9) & \cdot (1-0.95) & \cdot\ 0.1 & \cdot\ 0.2\ + & e \\ & & (1-0.8) & \cdot (1-0.9) & \cdot\ 0.1 & 0.2) & \neg e \end{array}$$

$$= 1.4 \cdot 10^{-5}$$

→

Dave Schultz

#1 Continued

$$P(d|b,\neg a,j,m) = \frac{P(d,b,\neg a,j,m)}{P(b,\neg a,j,m)} =$$

$$= \frac{P(d,b,\neg a,j,m)}{P(d,b,\neg a,j,m)+P(\neg d,b,\neg a,j,m)} = \frac{2.625\cdot 10^{-4}}{2.625\cdot 10^{-4}+1.4\cdot 10^{-5}}$$

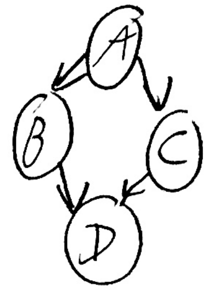
$$= \underline{\underline{0.9493670886}}$$

Lane Schultz

#2 a)

$$P(a) = \frac{True}{True + False} = \frac{270}{270+230} = \frac{270}{500} = 0.54$$

$$P(\neg a) = 1 - P(a) = 1 - 0.54 = 0.46$$



Can use a simmilar methodology for all

P(A)
<0.54, 0.46>

<P(True), 1 − P(True)>

For False values I use 1 − True

A	P(B|A)
T	$\left\langle \frac{122}{122+148}, \frac{148}{122+148} \right\rangle$
F	$\left\langle \frac{115}{115+115}, \frac{115}{115+115} \right\rangle$

→

A	P(B|A)
T	<0.452, 0.548>
F	<0.5, 0.5>

A	P(C|A)
T	$\left\langle \frac{138}{138+132}, \frac{132}{138+132} \right\rangle$
F	$\left\langle \frac{137}{137+93}, \frac{93}{137+93} \right\rangle$

→

A	P(C|A)
T	<0.511, 0.489>
F	<0.596, 0.404>

B	C	P(D|B, C)
T	T	$\left\langle \frac{66}{66+67}, \frac{67}{66+67} \right\rangle$
T	F	$\left\langle \frac{66}{66+38}, \frac{38}{66+38} \right\rangle$
F	T	$\left\langle \frac{64}{64+78}, \frac{78}{64+78} \right\rangle$
F	F	$\left\langle \frac{62}{62+59}, \frac{59}{62+59} \right.$

→

B	C	P(D|B, C)
T	T	<0.496, 0.504>
T	F	<0.635, 0.365>
F	T	<0.451, 0.549>
F	F	<0.512, 0.488>

→

[illegible]

#2 b)

A	B	C	D
T	F		T
T	F		T
T	F		T
T	F		T
T	F		T
T	F		T
T	F		T
T	F		T
T	F		T
T	F		T

Since all instances have $A=T$, $B=F$, $D=T$ we only need to compute $P(c \mid a, \neg b, d)$ and $P(\neg c \mid a, \neg b, d)$ from prior data

E-step

$$P(c \mid a, \neg b, d) = \frac{P(a, \neg b, c, d)}{P(a, \neg b, c, d) + P(a, \neg b, \neg c, d)}$$

$$= \frac{0.54 \cdot 0.548 \cdot 0.511 \cdot 0.451}{0.54 \cdot 0.548 \cdot 0.511 \cdot 0.451 + 0.54 \cdot 0.548 \cdot 0.489 \cdot 0.51}$$

$$= \underline{0.48}$$

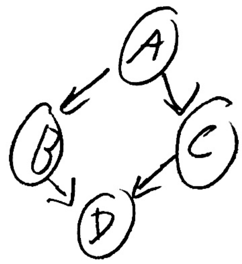
$$P(\neg c \mid a, \neg b, d) = 1 - P(c \mid a, \neg b, d) = \underline{0.52}$$

$\langle 0.48, 0.52 \rangle$ for the C column rows

the values will most likely be False

$\longrightarrow$

#2 b) continued



M-step

The new count for A is $\langle 280, 230 \rangle$

$$P(a) = \frac{280}{280+230} = 0.55$$

$$P(\neg a) = 1 - P(a) = 0.45$$

P(A)
$\langle 0.55, 0.45 \rangle$

A	B|A
T	$\langle 122, 158 \rangle$
F	$\langle 115, 115 \rangle$

$\longrightarrow$

A	P(B|A)
T	$\langle 0.44, 0.56 \rangle$
F	$\langle 0.5, 0.5 \rangle$

$$P(b|a) = \frac{122}{122+158} = 0.44$$

$$P(\neg b|a) = 1 - P(b|a) = 0.56$$

$$P(b|\neg a) = 1/2$$

$$P(\neg b|\neg a) = 1 - P(b|\neg a) = 1/2$$

Vance Schultz

#2 b) Continued

By maximum likelihood:

$$P(c|a) = \frac{E\#(a \wedge c)}{E\#(c)} = \frac{10 \cdot 0.48 + 138}{10 + 138 + 132} = 0.51$$

$$P(\neg c|a) = 1 - P(c|a) = 0.49$$

$$P(c|\neg a) = \frac{0 \cdot (0.48) + 137}{0 + 137 + 93} = 0.60$$

$$P(\neg c|\neg a) = 1 - P(c|\neg a) = 0.40$$

A	P(C|A)
T	⟨0.51, 0.49⟩
F	⟨0.60, 0.40⟩

→ Unchanged but the problem said use 2 sig figs.

→

Zane Schultz

#2 b) continued

B	C	P(D|B, C)
T	T	$\langle 0.50, 0.50 \rangle$
T	F	$\langle 0.63, 0.37 \rangle$
F	T	$\langle 0.45, 0.55 \rangle$
F	F	$\langle 0.51, 0.49 \rangle$

$$P(d \mid b, c) = \frac{E\#(b \wedge c \wedge d)}{E\#(b \wedge c)}$$

$$= \frac{66 + (0.48) \cdot 0}{66 + 67 + 0}$$

$= 0.50 \rightarrow$ Same as before because we only added False B values

$$P(\neg d \mid b, c) = 1 - P(d \mid b, c) = 0.50$$

$$P(d \mid \neg b, c) = \frac{64 + 10(0.48)}{64 + 78 + 10} = 0.45$$

$$P(\neg d \mid \neg b, c) = 0.55$$

$$P(d \mid b, \neg c) = \frac{66 + 0 \cdot (0.48)}{66 + 38 + 0} = 0.63$$

$$P(\neg d \mid b, \neg c) = 0.37$$

$$P(d \mid \neg b, \neg c) = \frac{62 + 0.48 \cdot 10}{62 + 59 + 10} = 0.51$$

$$P(\neg d \mid \neg b, c) = 0.49$$