CS 228: Probabilistic Graphical Models

Luke Jaffe

Due: 01/24/17
Submitted: 01/26/17

(using 2 late days)

Problem 1: Probability Theory

Let D = Have Disease
Let T = Test Positive.

Given:
$P(T|D) = 0.99$
$P(\bar{T}|\bar{D}) = 0.99$
$P(D) = 0.0001$

Question:
what is $P(D|T)$?

Since the disease is rare, the chances X actually has it are quite low.

Using Bayes' Rule:

$P(D|T) = \frac{P(T|D)P(D)}{P(T)}$, missing $P(T)$

We also have $P(\bar{T}|D) = 1 - P(T|D)$
and $P(T|\bar{D}) = 1 - P(\bar{T}|\bar{D})$

Using Bayes' Rule:

$P(\bar{D}|T) = \frac{P(T|\bar{D})P(\bar{D})}{P(T)}$, $P(D|T) = \frac{P(T|D)P(D)}{P(T)}$

$P(D|T) = 1 - P(\bar{D}|T) \rightarrow 1 - \frac{P(T|\bar{D})P(\bar{D})}{P(T)} = \frac{P(T|D)P(D)}{P(T)}$

$\rightarrow P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D}) = 0.99 \cdot 0.0001 + 0.01 \cdot 0.9999 \approx 0.0101$

$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{0.99 \cdot 0.0001}{0.0101} = 0.0098$

$\boxed{P(D|T) = 0.0098}$

Problem 2: Review of Dynamic Programming

- Give an $O(m^2n)$ algorithm for solving: $\max_{x_1,\dots,x_n \in S} P(x_1,\dots,x_n)$

Nothing depends on X_n besides the last term $P(X_n|X_{n-1})$, so the only m entries we care about in the last table are the m entries where $\max_{X_n} P(X_n|X_{n-1})$ for all $X_{n-1} \in S$. This takes $O(m^2)$ to find and store these entries

Now that we have calculated the the max of the last term for all values of $X_{n-1} \in S$, let's look at term $n-1$: $P(X_{n-1}|X_{n-2})$. Since we already know which m entries of the last table we will use, we just need to use these values to calculate the maximum product of $P(X_{n-1}|X_{n-2})P(X_n|X_{n-1})$ for all possible values of $X_{n-1} \in S$

Pseudocode: Runtime analysis next to each line

```
max_arr = m zeros
For each v_i in S:                                  O(m)
    v_i_arr = m zeros
    For each v_j in S:                              O(m)
        v_i_arr[j] = P(x_{n-1} = v_j | x_{n-2} = v_i) · max_{x_n} P(x_n | x_{n-1} = v_j)   O(1)
    max_arr[i] = max(v_i_arr)                       O(m)
```

Runtime of loops: $O(m \cdot (m \cdot 1 + m)) = O(m^2)$

Now, iterate backwards through all remaining tables, repeating this process, and storing the results for use by the next computation. There will be $n-2$ iterations before reaching X_1, for $O(m^2 n)$ time

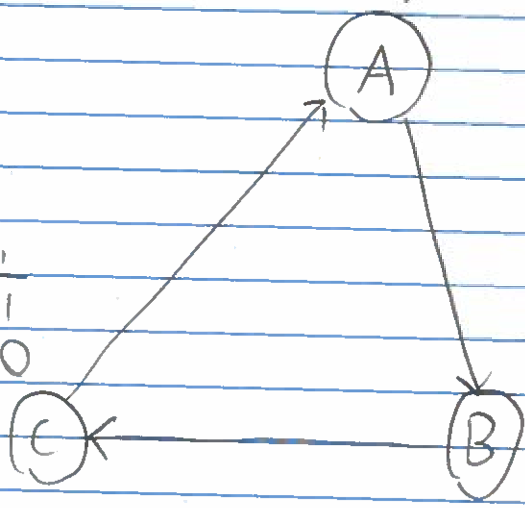
Finally, after reaching X_1, do a pass back through all the stored tables, following the route for each $X_1 \in S$. This pass takes $O(mn)$ time.

Final run time $= O(m^2n) + O(mn) = \boxed{O(m^2n)}$

Problem 3: Bayesian Networks

Observe the following Bayes' net and associated probability tables:

This graph has a directed cycle



	a_0	a_1
c_0	0	1
c_1	1	0

All tables are valid, with marginals summing to 1

	c_0	c_1
b_0	0	1
b_1	1	0

	b_0	b_1
a_0	0	1
a_1	1	0

Now, we will compute $P(A,B,C)$ with the following table:

a	b	c	$P(A=a, B=b, C=c)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	0

Sum = 0

Since the joint probability $P(A,B,C)$ does not $= 1$, this network defines an invalid probability distribution. Thus, we have proved by example that if G has a directed cycle, f may no longer define a valid probability distributions.

lem 4: Conditional Independence

1. $P(B,\eta)$, $P(a)$, $P(B|a)$, $P(\eta|a)$

$P(a|B,\eta) = \frac{P(B,\eta|a)P(a)}{P(B,\eta)}$ → $P(a)$ ✓, $P(B,\eta)$ ✓, $P(B,\eta|a)$ X

$\boxed{NO}$ we need to know $B \perp \eta | a$ to compute $P(B,\eta|a)$ using $P(B|a)$, $P(\eta|a)$

✓2. $P(B,\eta)$, $P(a)$, $P(B,\eta|a)$

$P(a|B,\eta) = \frac{P(B,\eta|a)P(a)}{P(B,\eta)}$ → $P(B,\eta|a)$ ✓, $P(a)$ ✓, $P(B,\eta)$ ✓ $\boxed{YES}$

X3. $P(B|a)$, $P(\eta|a)$, $P(a)$

$\boxed{NO}$ This has strictly less information than 1., so it is also insufficient

✓1. If $B \perp \eta | a$, then $P(B,\eta|a) = P(B|a) \cdot P(\eta|a)$.

Now we have enough info to compute $P(a|B,\eta)$ $\boxed{YES}$

✓2. $\boxed{YES}$ We have more information, so still yes.

X3. We can now compute $P(B,\eta|a) = P(B|a) \cdot P(\eta|a)$
But we still cannot compute $P(B,\eta)$, so there is still insufficient information. $\boxed{NO}$

Problem 5: Bayesian Networks

1.

a. First, $A \perp B$, from graph structure, d-sep$_B$(A;B) = True

Proof: - $A \to D \leftarrow B$ is blocked by V-structure

- $A \to C \to F \to H \leftarrow E \leftarrow B$ is blocked by $F \to H \leftarrow E$

$P(A=0, B=0) = 0.8 \cdot 0.3 = \boxed{0.24}$

b. Since d-sep$_B$(A;E) = True, $A \perp E$

Proof: - $E \leftarrow B \to D \leftarrow A$ blocked by $B \to D \leftarrow A$ V-structure

- $E \leftarrow B \to D \to F \leftarrow C \leftarrow A$ blocked by $D \to F \leftarrow C$

- $E \to H \leftarrow F \ldots$ blocked by V structure

$P(E=1 \mid A=1) = P(E=1 \mid B) = P(E=1 \mid B=0) \cdot P(B=0) + P(E=1 \mid B=1) \cdot P(B=1)$

$= 0.9 \cdot 0.3 + 0.1 \cdot 0.7 = \boxed{0.34}$

2.

a. d-sep$_B$(A;E | {D,H}) = $\boxed{\text{False}}$

- $A \to C \to F \to H \leftarrow E$ is an active path

$A \to C \to F$ = open

$C \to F \to H$ = open

$F \to H \leftarrow E$ = open, V-structure w/ middle given

b. d-sep$_B$(G;E|D) = $\boxed{\text{False}}$

- $G \leftarrow F \leftarrow C \leftarrow A \to D \leftarrow B \to E$ is an active path

$G \leftarrow F \leftarrow C$ = open

$F \leftarrow C \leftarrow A$ = open

$C \leftarrow A \to D$ = open

$A \to D \leftarrow B$ = open, V-structure w/ middle given

$D \leftarrow B \to E$ = open

c. d-sep$_B$({A;B}, {G;H} | F) = $\boxed{\text{False}}$

- all pairs {A;G}, {A;H}, {B;G}, {B;H} must be d-sep

- $B \to E \to H$ is an active path, just takes one active path for one pair to make statement false

Problem 6: Bayesian Networks and Explaining Away

1. Expected creativity of a student

$$E[C] = \int_c c\,dc = \int_0^1 c\,dc = \frac{c^2}{2}\Big|_0^1 = \frac{1}{2} \qquad \boxed{E[C] = \frac{1}{2}}$$

2. Expected creativity of admitted students

$$E[C \mid A=1] = \int_c c\,P(C \mid A=1)\,dc = \int_c c\left[\int_I P(C, I \mid A=1)\,dI\right] dc$$

$$= \int_{0.5}^{1} c\left[\int_{1.5-c}^{1} P(C \mid A=1)\,di\right] dc = \boxed{\frac{5}{6}}$$

3. Expected creativity of student w/ $I = 0.95$

Since $C \perp I$, $E[C \mid I = 0.95] = E[C] = \boxed{\frac{1}{2}}$

4. $E[C \mid A=1, I=0.95] = \int_c c\,P(C \mid A=1, I=0.95)\,dc = \boxed{0.775}$

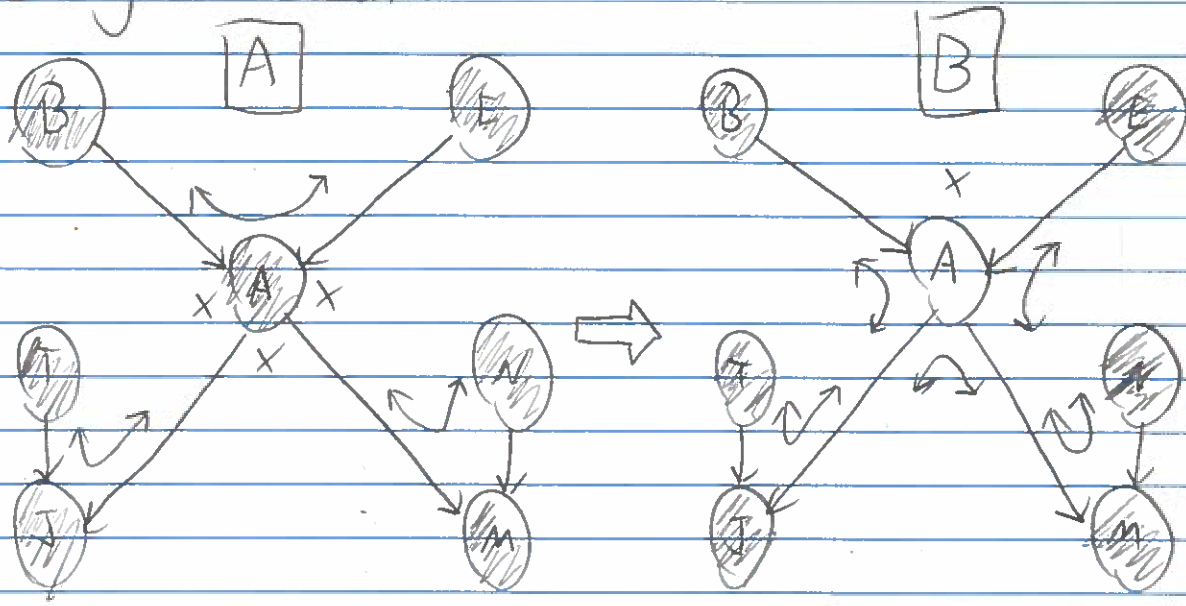
From python program

Not sure about answers to 2, 4, verified using attached python script.

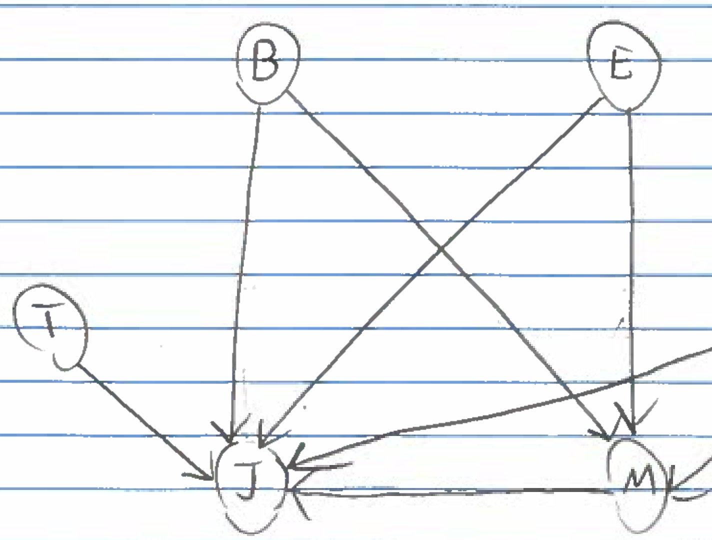
```
“””
Python script to check for problem 6
“””

#!/usr/bin/env python

import numpy as np

n = 100000000

c = np.random.rand(n)
i = np.random.rand(n)

# Calculate E(C)
print "1. E(C) = {}".format(c.mean())

# Calculate P(A=1)
s = c+i
print "P(A=1) = {}".format(len((s>1.5).nonzero()[0])/float(n))

# Calculate E(C|A=1)
aidx = s>1.5
print "2. E(C|A=1) = {}".format(c[aidx].mean())
print "P(C|A=1) = {}".format(len(c[aidx])/float(n))

# Calculate P(C|I=0.95)
lb,ub = 0.95-0.00005, 0.95+0.00005
lidx = (i>lb).nonzero()
uidx = (i<ub).nonzero()
iidx = lidx[0][np.in1d(lidx, uidx)]
print "3. E(C|I=0.95) = {}".format(c[iidx].mean())

# Calculate P(C|A=1, I=0.95)
aidx = aidx.nonzero()[0]
idx = iidx[np.in1d(iidx, aidx)]
print "4. E(C|A=1,I=0.95) = {}".format(c[idx].mean())
```

Problem 7: Bayesian Networks

1. By marginalizing out Alarm we are treating it like it's not observed, and examining conditional independencies that hold in the original network.



Once alarm is no longer observed, "new" dependencies are created. We can check for this w/ d-separability. To maintain minimality of the I-map, we must make sure there is exactly one active path between any dependent nodes in the new graph. Here is one possible minimal I-map:



We need active paths between all nodes w/ active paths in graph B above. This is one solution.

Problem 7. Bayesian Networks

2. To generalize the procedure:

- Treat the removed node of the graph as unobserved, and the other nodes as observed.
- Analyze the d-separability for all pairs of nodes in this graph with the above information.
- Now, remove the targeted node from the graph, and all edges from that node, leaving all other edges intact.
- Iterate through all the d-separability analysis for every node pair from the previous step
- If the previous graph had an active path for two nodes that is not present in the node-removed graph, add an edge to create that path
- Then, check against all node pairs to see if needed active paths have been created, marking them if they have
- Repeat this process of adding edges to create active paths that are in the "unobserved" graph, till all dependency information has been restored.
- This graph is now a minimal I-map of $P'_{BN}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$

Problem 8: Towards inference in Bayesian Networks

1. This problem can be solved with a recursive algorithm:

function Compute(X_i):
    If X_i has no parents:
        return $P(X_i)$
    Else:
        For all Parents of X_i, store Compute($Pa(X_i)$)
        return $P(X_i \mid Pa(X_i))$

2. For any joint distribution $P(X,Y)$ you can sample by first drawing a sample $x \sim P(X)$, then drawing $y \sim P(Y \mid X=x)$. This follows from the chain rule: $P(X,Y) = P(X)\,P(Y \mid X)$

Algorithm:

- First, draw a sample $x_i \sim P(X_i)$ for all root nodes of the network. mark these nodes as covered
- Then, for all children of covered nodes, draw a sample from any nodes that depend solely on previously covered nodes using the random sample multinomial function guaranteed by the problem, based on the assertion above. cover these sampled nodes.
- Repeat this process till all nodes in the graph have been covered, and you will have a sample for each parameter.

Problem 9: Programming Assignment

1. There are $2^{28\times 28} = 2^{784}$ binary images.

2. You would need $2^{28\times 28} - 1 = 2^{784} - 1$ parameters to specify an arbitrary distribution over all 28×28 binary images.

3. The joint pdf for the Bayes' net in figure 1 can be written as:

$$P(z_1, z_2, x_{1:784}) = P(z_1)P(z_2)\cdot P(x_1|z_1,z_2)\cdot\ldots\cdot P(x_{784}|z_1,z_2)$$

The conditional terms have 3 parameters each, so

$$\#\text{ Parameters} = 784\cdot 3 + 2 = \boxed{2354}$$

5. The intuitive role of the z_1, z_2 parameters in this model is as prior beliefs for the model.

7. The relationship between the images is that they correspond spatially, except flipped over the line $y = x$