CS 181 LECTURE 4/2/24

Scribe Notes

Contents

1 Where We Are in the Course

Today we are focusing on inference. We are continuing the topic of structure in models.

2 Real World Story

There are different ways of writing down structure in probability distributions.

Frances Haugen (Facebook whistleblower and author of The Power of One) exposed a lot of problems. To what extent was Facebook responsible for mental health problems among teenagers? There are a lot of people trying to get data from Facebook, as well. So, one question that was investigated is how do you catch a bot trying to exploit data?

We can imagine, there are a lot of people in Facebook and we can draw lines between "friends."

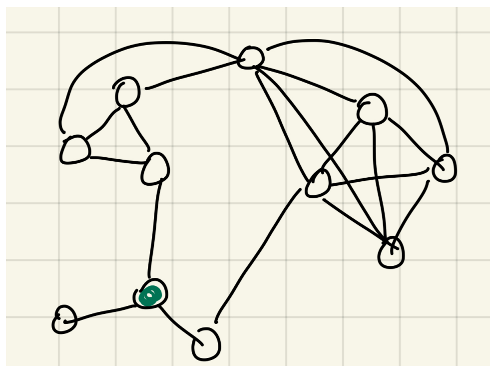


Figure 1: Friendships on Facebook.

There are groups of people who know each other well.

Condition on the fact that Finale always look at her own profile, which other profiles can she see?

At some point, the connections will be too far from her profile where she is very unlikely to be looking at a profile. So, there is a distribution that we can draw, given that we start at Finale's profile. This is a distribution of actual use. People who are trying to hack into Facebook and scrape data will not have access to this distribution. So, we may have odd looking distributions - like a person accessing profiles in very different communities - is likely a bot and not a real person browsing Facebook.

This is an example of how characterizing the structure of a distribution is practical - in this case, identifying bots that are causing chaos on a platform.

3 Marginalizing Data

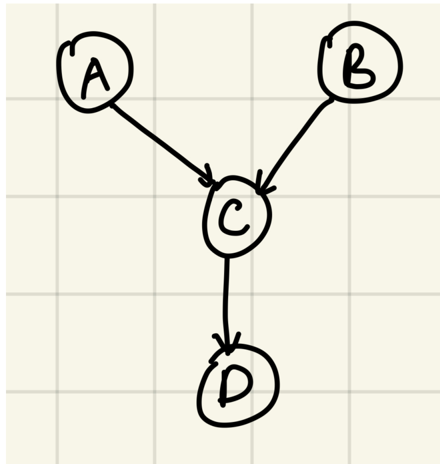


Figure 2: Example DAG.

$$p(A, B, C, D) = p(A)p(B)p(C|A, B)p(D|C)$$

We only care about one variable. So we marginalize.

$$p(D) = \sum_A \sum_B \sum_C p(A, B, C, D)$$

Let each node be discrete with k-values. So, $p(A, B, C, D)$ is a k^4-sized object. $p(D)$ is a k-sized object. The whole idea is that k^4 is pretty big - and if we had more nodes, then the object would be even bigger. We want a way to get to $p(D)$ without having to think about the huge object. The key is noting that the big object is made up of smaller objects. For example, $p(A)$ is size k, $p(B)$ is also size k.

Let's go through a simple example.

3.1 Concrete Example

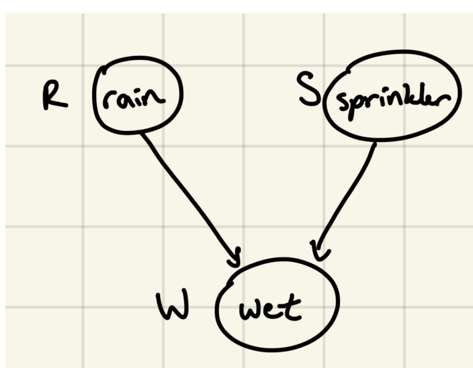


Figure 3: Wet lawn.

Let's say $p(S) = 1/2$ and $p(R) = 1/4$.

We want to calculate the probability that the lawn is wet.

S	R	p(W|S,R)
T	T	99/100
T	F	9/10
F	T	9/10
F	F	0

Figure 4: Table of probabilities.

Now, we will try to do some inference.

$$p(\text{query}|\text{evidence})$$

We will use Bayes rule.

$$p(R=1|W=1) = \frac{p(W=1,R=1)}{p(R=1,W=1)+p(R=0,W=1)}$$
$$= \frac{p(W=1,R=1,S=0)+p(W=1,R=1,S=1)}{p(R=1,W=1,S=0)+p(R=1,W=1,S=1)+p(R=0,W=1,S=0)+p(R=0,W=1,S=1)}$$

In the second step, we are marginalizing over the possible values of S.

Note: $p(W=1,R=1,S=1) = p(S=1)p(R=1)p(W=1|S=1,R=1)$ based on the structure of our Bayes net, depicted in Fig 3.

Now, we can plug in the numbers for our probabilities.

$$p(R=1|W=1) = \frac{p(W=1,R=1)}{p(R=1,W=1)+p(R=0,W=1)}$$
$$= \frac{1/4 \cdot 1/2 \cdot 9/10 + 1/4 \cdot 1/2 \cdot 99/100}{1/4 \cdot 1/2 \cdot 9/10 + 1/4 \cdot 1/2 \cdot 99/100 + 3/4 \cdot 1/2 \cdot 0 + 3/4 \cdot 1/2 \cdot 9/10}$$
$$= 21/51 = 0.4$$

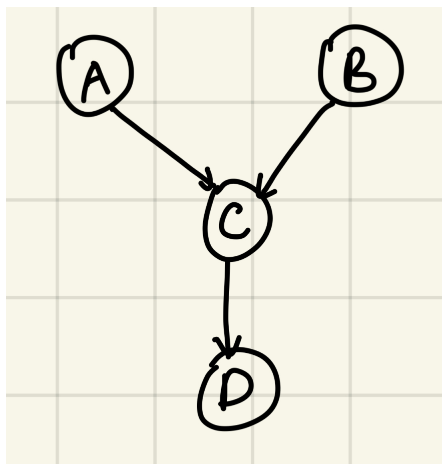
Compare to: $p(R=1) = \frac{1}{4} = 0.25$

So, we see that knowing that the lawn in wet increases the probability that there was rain. Suppose now that we knew that the sprinkler was one and the lawn is wet.

$$
\begin{aligned}
p(R=1|S=1,W=1) &= \frac{p(W=1,S=1,R=1)}{p(R=1,W=1,S=1)+p(R=0,W=1,S=1)} \\
&= 11/41 = 0.26
\end{aligned}
$$

When we first stepped out and we saw that the lawn was wet, we would have thought the probability that there was rain was at 40 percent. But then we find out that the sprinkler was turned out and the probability that there was rain goes back down to around 26 percent. So, observing a variable changed the probabilities of the other variables. If we observe the wetness of the lawn, then knowing something about the sprinkler would tell us something about the rain.

3.2 Inference In General



In our previous example, we were writing out all the probabilities. With a bigger graph, the sums get really long.

$$
\begin{aligned}
p(D) &= \sum_A \sum_B \sum_C p(A,B,C,D) \\
&= \sum_A \sum_B \sum_C p(A)p(B)p(C|A,B)p(D|C)
\end{aligned}
$$

What is the order of summing that we should do? Let's say we sum over C first.

$$
= \sum_C p(D|C) \sum_B p(B) \sum_A p(A)p(C|A,B)
$$

We know that $p(C|A,B)$ is a k^3-sized object because for every value of A and B there is a probability for every value of C. And $p(A)$ has size k. So, $\sum_A p(A)p(C|A,B)$ has size k^2.

Then, after we sum over B, we have $g(C)$ which is a size k-object. In other words, $\sum_B p(B) \sum_A p(A)p(C|A,B)$ is a k-sized object.

Furthermore, $p(D|C)$ is a k^2-sized object and then summing over C gives us a final object of size k.

If we have ordered our sum in a different way, then we could end up needing to use larger objects.

For example, say we have $\sum_A p(A) \sum_B p(B) \sum_C p(C|A,B)p(D|C)$, then $\sum_C p(C|A,B)p(D|C)$ is a k^4-sized object.

Thus, the order of elimination can change the size of the objects and computations that we are dealing with.

4 Order of Elimination

What order should we choose?

Unfortunately, in general, this is a NP-hard problem. But, for a specific class of problems, called polytrees, there is a simple strategy.

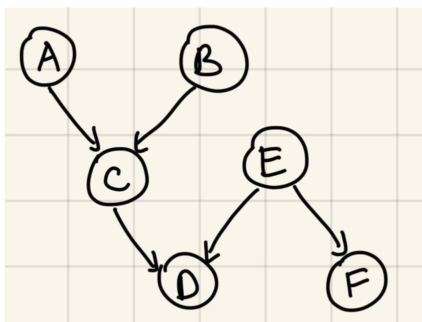


Figure 5: Example of polytree.

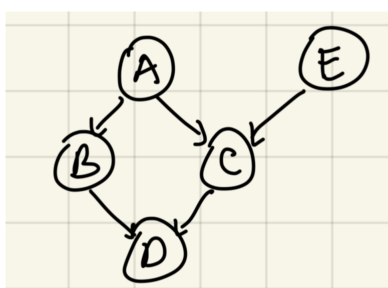


Figure 6: Example of not a polytree.

Fig 5 is an example of a polytree. Fig 6 is not a polytree because there is an undirected loop between A, B, C, D.

4.1 Strategy for Polytrees

1) Prune variables that are not ancestors of Q or E.

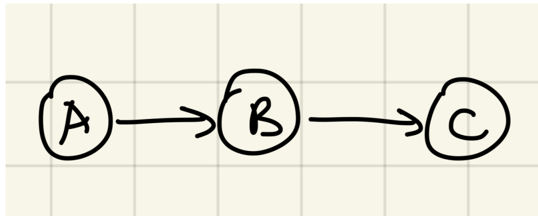


Figure 7: Example polytree.

If we are trying to find $p(B|A = 1)$, we can completely ignore C because it is not an ancestor of A or B.
2) Find the leaves and work backwards.

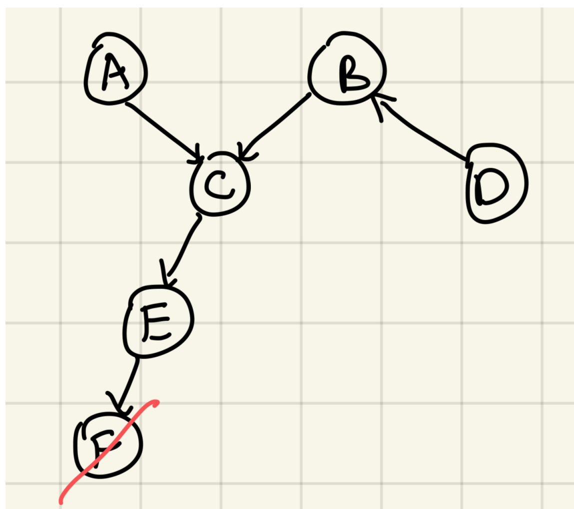


Figure 8: Another example.

Say we are trying to find $p(E|D = 1)$. We know we can ignore F by step 1. Then, a possible choice of ordering is A, B, C or B, A, C. We know we want to do A and B before C because C is closer to E.

Let us apply this on another example.

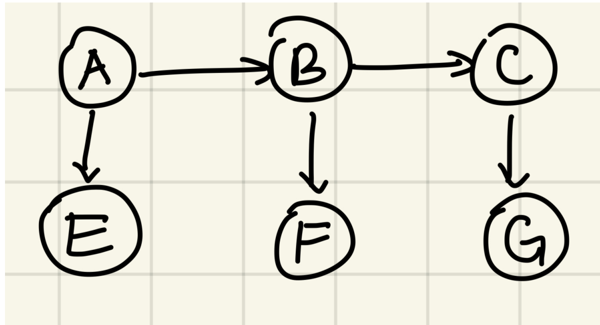


Figure 9: Another example.

This is the model for a Hidden Markov Model. At every point in time, we can deduce some information. For example, we could be looking at health data.

Suppose we want $p(C|E,F,G)$. For example, say we have the GPS info for the last three hours and we want to know where we are now.

$$p(C|E,F,G) = \sum_A \sum_B p(A)p(B|A)p(E|A)p(F|B)p(C|B)p(G|C)$$

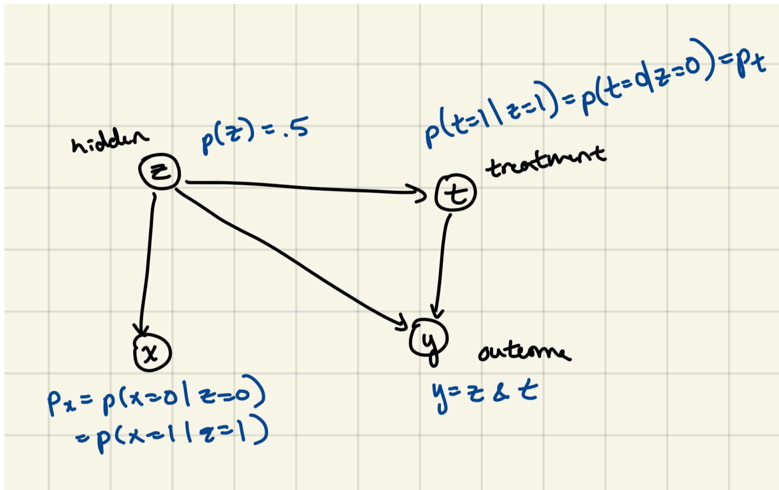
So we need to decide whether we sum A first or B first. In this case, if we root at C, then B is the closest out of B and A. So, we want to leave B for last.

$$= p(G|C) \sum_B p(F|B)p(C|B) \sum_A p(A)p(B|A)p(E|A)$$

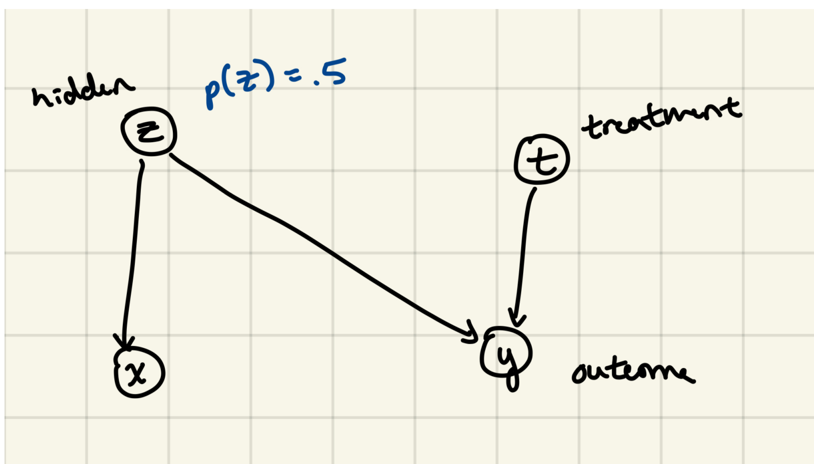
Then, $\sum_A p(A)p(B|A)p(E|A)$ turns into a $g(B)$. Note that we are not pruning G because G is part of the evidence.

5 Concept Check

Suppose we do $t = 1$, which means that we intervene with treatment.



What part of the graph changes?
There is no longer an arrow from z to t, because t is being set to 1 by us.



What is $p(y = 1|\text{do } t = 1)$? Answer: $p(y = 1|\text{do } t = 1) = 0.5$
Compute $p(y = 1|\text{do } t = 1) = \sum_x p(y = 1|t = 1, x = x)p(x)$.
Fill out the following table.

z	x	t	y	prob
0	0	0	0	?

6 Concept Check Solution

z	x	t	y	prob
0	0	0	0	
0	1	1	1	
1	1	1	1	
0	0	1	1	
1	0	1	1	

(The table isn't filled out completely but feel free to do on your own!)

$$
\begin{aligned}
p(y=1|\text{do } t=1) &= \sum_x p(y=1|t=1, x=x)p(x) \\
&= p(y=1|t=1, x=1)p(x=1) + p(y=1|t=1, x=0)p(x=0)
\end{aligned}
$$

This concept check shows us that if you mistakenly believe that z does not exist, you would have an incorrect result. Therefore, knowing the structure of a model is extremely important.