Foundations of Machine Learning: Week 1: Bayesian Networks

Professor Christopher Yau
christopher.yau@manchester.ac.uk

October 21, 2020

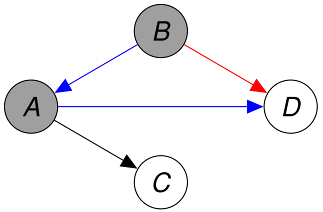
Foundations of Machine Learning:Week 1: Bayesian Networks

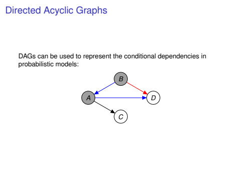
Foundations of Machine Learning:
Week 1: Bayesian Networks

Professor Christopher Yau
christopher.yau@manchester.ac.uk
October 21, 2020

Directed Acyclic Graphs

DAGs can be used to represent the conditional dependencies in probabilistic models:



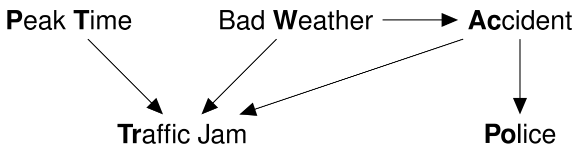


In the previous lecture, we saw how directed acyclic graphs can be used to represent complex conditional dependencies between random variables in a probability model.

In this lecture, we are now going to see how this can be used in numerical examples for problems involving Bayesian Networks.

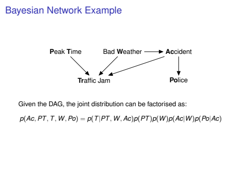
In these examples we will only look at binary random variables but principles remain the same for non-binary variables.

Bayesian Network Example



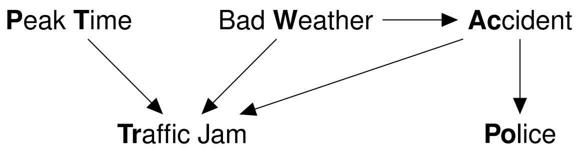
Given the DAG, the joint distribution can be factorised as:

$$p(Ac, PT, T, W, Po) = p(T|PT, W, Ac)p(PT)p(W)p(Ac|W)p(Po|Ac)$$



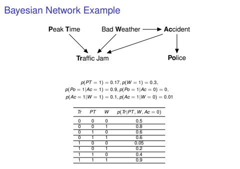
Here is a DAG for an example Bayesian network problem.
The probability of a traffic jam depends on whether it is peak time, there is bad weather, or a road accident.
Accidents are affected by bad weather.
And accidents affect the probability that police assistance would be required.
Given the DAG, the joint distribution can be factorised with the form shown.

Bayesian Network Example



$$p(PT=1)=0.17, p(W=1)=0.3,$$
$$p(Po=1|Ac=1)=0.9, p(Po=1|Ac=0)=0,$$
$$p(Ac=1|W=1)=0.1, p(Ac=1|W=0)=0.01$$

Tr	PT	W	$p(Tr\|PT,W,Ac=0)$
0	0	0	0.5
0	0	1	0.8
0	1	0	0.6
0	1	1	0.6
1	0	0	0.05
1	0	1	0.2
1	1	0	0.4
1	1	1	0.9



Now suppose we assign some actual numerical quantities to these probability distributions.

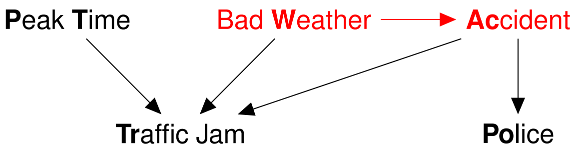
The model is quite complex and there are many probabilities and combinations to account for.

If we do any computations, we want to avoid having to use these probabilities unnecessarily.

For example, the conditional probability of a traffic jam depends on three variables and so there are $2^3 = 8$ combinations on inputs into the conditional probability.

Total Probability

Question: Compute the total probability of an accident: $p(Ac = 1)$.



There is only one path leading to an accident.

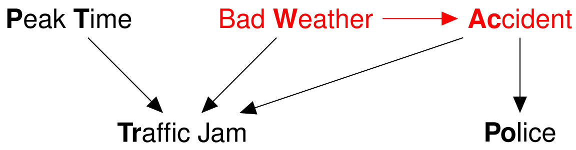
$$p(Ac = 1) = \sum_{W=0}^{1} p(Ac = 1|W)p(W)$$

Total Probability
Question: Compute the total probability of an accident: $p(Ac = 1)$.
There is only one path leading to an accident.
$p(Ac = 1) = \sum_{W=0}^{1} p(Ac = 1|W)p(W)$

Lets compute the total probability of an accident.
First, recognise that there is only one path into an accident (Ac) from bad weather (W).
So, from the DAG, the total probability of an accident has to sum out the probability of bad weather (W) since bad weather is a parent of accident (Ac).

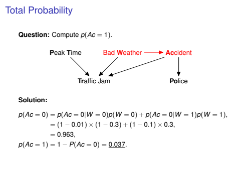
Total Probability

Question: Compute $p(Ac = 1)$.



Solution:

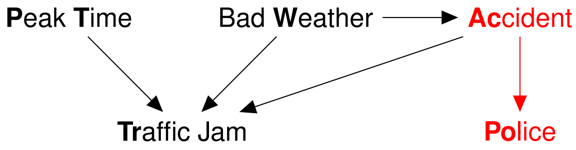
$$
\begin{aligned}
p(Ac = 0) &= p(Ac = 0|W = 0)p(W = 0) + p(Ac = 0|W = 1)p(W = 1),\\
&= (1 - 0.01) \times (1 - 0.3) + (1 - 0.1) \times 0.3,\\
&= 0.963,\\
p(Ac = 1) &= 1 - P(Ac = 0) = \underline{0.037}.
\end{aligned}
$$



Adding in our numerical quantities, we can now compute the total probability of an accident or it not occurring and then using the complement law.
The key thing is to recognise from the DAG that to answer *this* question, you did not have to use the full joint distribution and every numerical quantity given to you.
The only thing that accidents depend on is bad weather and as bad weather does not depend on anything else we only need to look at this connection.

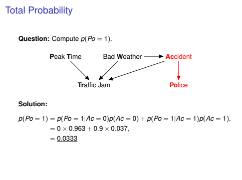
Total Probability

Question: Compute $p(Po = 1)$.

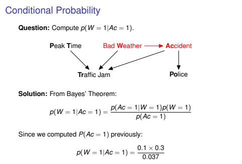


Solution:

$$\begin{aligned} p(Po = 1) &= p(Po = 1|Ac = 0)p(Ac = 0) + p(Po = 1|Ac = 1)p(Ac = 1), \\ &= 0 \times 0.963 + 0.9 \times 0.037, \\ &= \underline{0.0333} \end{aligned}$$

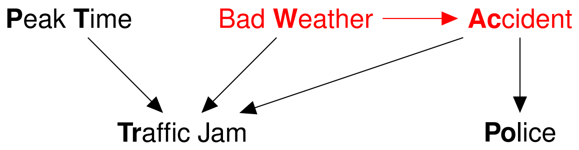


Now, having computed the total probability of an accident, if I wanted to compute the total probability that the police are called to an accident, things are simplified.
Since the probability that the police are called depends only on whether an accident has occurred, we can re-use the total probability of an accident that we previously calculated in the usual total probability formula.
Although bad weather is on the path that leads to the police being called, we have already taken its influence into account when computing the total probability of an accident. Since there is no other path into police from bad weather that variable can be eliminated if we know the total probability of an accident.
Recognising these connections will save you from having to re-derive probabilistic relationships from first principles each time.


Conditional Probability
Question: Compute $p(W = 1|Ac = 1)$.
Peak Time
Bad Weather
Accident
Traffic Jam
Police
Solution: From Bayes' Theorem:
$p(W = 1|Ac = 1) = \frac{p(Ac = 1|W = 1)p(W = 1)}{p(Ac = 1)}$
Since we computed $P(Ac = 1)$ previously:
$p(W = 1|Ac = 1) = \frac{0.1 \times 0.3}{0.037}$

Conditional Probability

Question: Compute $p(W = 1|Ac = 1)$.



Solution: From Bayes' Theorem:

$$p(W = 1|Ac = 1) = \frac{p(Ac = 1|W = 1)p(W = 1)}{p(Ac = 1)}$$

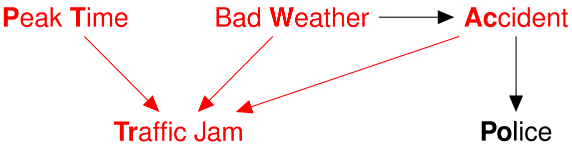
Since we computed $P(Ac = 1)$ previously:

$$p(W = 1|Ac = 1) = \frac{0.1 \times 0.3}{0.037}$$

Now suppose we are interested in the conditional probability of bad weather given there was an accident.
From Bayes' Theorem. we can rewrite this conditional probability.
The key thing to note is the denominator involving the total probability of an accident is something we have computed already.
So we can therefore apply Bayes' Theorem with ease.
If we hadn't computed the total probability of an accident previously, we know how to do it and that process needn't involving summing over every variable - we can use the DAG to focus our efforts.

Conditional Probability

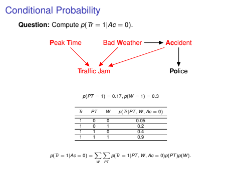
Question: Compute $p(Tr = 1|Ac = 0)$.



$$p(PT = 1) = 0.17, p(W = 1) = 0.3$$

Tr	PT	W	$p(Tr\|PT, W, Ac = 0)$
1	0	0	0.05
1	0	1	0.2
1	1	0	0.4
1	1	1	0.9

$$p(Tr = 1|Ac = 0) = \sum_{W} \sum_{PT} p(Tr = 1|PT, W, Ac = 0)p(PT)p(W).$$



Now, suppose we are interested in the conditional probability of a traffic jam given there have been no accidents.
If we examine the DAG, the set of variables and connections we need to consider are highlighted here in red. We therefore can ignore police.
Since, we are interested in the probability of a traffic jam, we need only consider those rows of the conditional probability of a traffic jam where the variable Tr is 1.
Further, since we are conditioning on $Ac = 0$, this means that we need only need to sum over peak time and weather and do not need to consider how accidents depend on weather.

Conditional Probability

Question: Compute $p(Tr = 1 | Ac = 0)$.

$$p(PT = 1) = 0.17, p(W = 1) = 0.3$$

Tr	PT	W	$p(Tr\|PT, W, Ac = 0)$
1	0	0	0.05
1	0	1	0.2
1	1	0	0.4
1	1	1	0.9

$$\begin{aligned} p(Tr = 1 | Ac = 0) &= \sum_{W} \sum_{PT} p(Tr = 1 | PT, W, Ac = 0) p(PT) p(W), \\ &= 0.05(1 - 0.17)(1 - 0.3) \\ &\quad + 0.2(1 - 0.17)(0.3) \\ &\quad + 0.4(0.17)(1 - 0.3) \\ &\quad + 0.9(0.17)(0.3) \end{aligned}$$

Conditional Probability
Question: Compute $p(Tr = 1 | Ac = 0)$.
$p(PT = 1) = 0.17, p(W = 1) = 0.3$
$p(Tr = 1 | Ac = 0) = \sum_{W} \sum_{PT} p(Tr = 1 | PT, W, Ac = 0) p(PT) p(W)$, $= 0.05(1 - 0.17)(1 - 0.3) + 0.2(1 - 0.17)(0.3) + 0.4(0.17)(1 - 0.3) + 0.9(0.17)(0.3)$

Using the numerical data, we can then compute the desired conditional probability without enumerating all the possibilities over all the random variables. The DAG saves us some work.

Recap

In this lecture, we applied the principles of marginalisation and conditional probability computations to a Bayesian Network example:

- Use the DAG to simplify the problem to consider only the elements which are *relevant* to the problem,
- Re-use computed quantities to answer related questions and avoid having to recompute from first principles each time.

Recap

In this lecture, we applied the principles of marginalisation and conditional probability computations to a Bayesian Network example:

- Use the DAG to simplify the problem to consider only the elements which are *relevant* to the problem,
- Re-use computed quantities to answer related questions and avoid having to recompute from first principles each time.

In this lecture, we applied the principles of marginalisation and conditional probability computations to a Bayesian Network example.
We saw how using DAGs can help us to simplify a complex probability problem by allowing us to focus only the elements which are *relevant* to the problem.
We also saw that we can re-use computed quantities to answer related questions and avoid having to recompute from first principles each time.

END LECTURE

END LECTURE