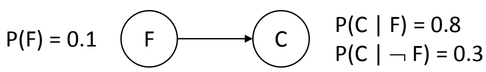
CS360 Homework 8– Solution

Bayesian Networks

1) Consider the following Bayesian network, where F = having the flu and C = coughing:

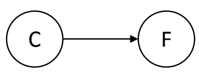


a) Write down the joint probability table specified by the Bayesian network.

Answer:

F	C	
t	t	$0.1 \times 0.8 = 0.08$
t	f	$0.1 \times 0.2 = 0.02$
f	t	$0.9 \times 0.3 = 0.27$
f	f	$0.9 \times 0.7 = 0.63$

b) Determine the probabilities for the following Bayesian network



so that it specifies the same joint probabilities as the given one.

Answer:
$P(C) = 0.08 + 0.27 = 0.35$

$P(F \mid C) = P(F, C) \;/\; P(C) = 0.08/0.35 \sim 0.23$

$P(F \mid \neg C) = P(F, \neg C) \;/\; P(\neg C) = 0.02/0.65 \sim 0.03$

c) Which Bayesian network would you have specified using the rules learned in class?

Answer:
The first one. It is good practice to add nodes that correspond to causes before nodes that correspond to their effects.

d) Are C and F independent in the given Bayesian network?

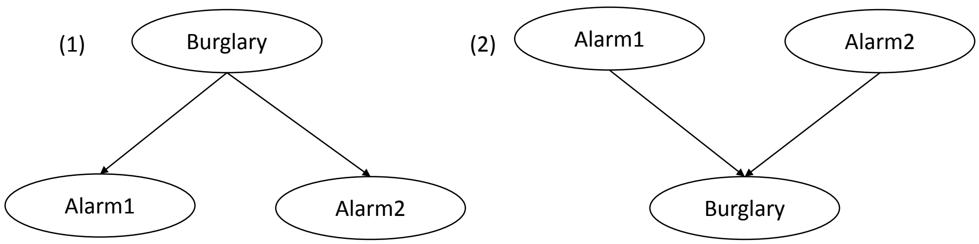
Answer:
No, since (for example) $P(F) = 0.1$ but $P(F \mid C) \sim 0.23$

e) Are C and F independent in the Bayesian network from Question b?

Answer:
No, for the same reason.

2) To safeguard your house, you recently installed two different alarm systems by two different reputable manufacturers that use completely different sensors for their alarm systems.

a) Which one of the two Bayesian networks given below makes independence assumptions that are not true? Explain all of your reasoning. Alarm1 means that the first alarm system rings, Alarm2 means that the second alarm system rings, and Burglary means that a burglary is in progress.



Answer:
The second one falsely assumes that Alarm1 and Alarm2 are independent if the value of Burglary is unknown. However, if the alarms are working as intended, it should be more likely that Alarm1 rings if Alarm2 rings (that is, they should not be independent).

b) Consider the first Bayesian network. How many probabilities need to be specified for its conditional probability tables? How many probabilities would need to be given if the same joint probability distribution were specified in a joint probability table?

Answer:
We need to specify 5 probabilities, namely P(Burglary), P(Alarm1 | Burglary), P(Alarm1 | ¬ Burglary), P(Alarm2 | Burglary) and P(Alarm2 | ¬ Burglary). A joint probability table would need $2^3 - 1 = 7$ probabilities.

c) Consider the second Bayesian network. Assume that:

P(Alarm1) = 0.1

P(Alarm2) = 0.2

P(Burglary | Alarm1, Alarm2) = 0.8

P(Burglary | Alarm1, ¬ Alarm2) = 0.7

P(Burglary | ¬ Alarm1, Alarm2) = 0.6

P(Burglary | ¬ Alarm1, ¬ Alarm2) = 0.5

Calculate P(Alarm2 | Burglary, Alarm1). Show all of your reasoning.

Answer:
P(Alarm2 | Burglary, Alarm1) = P(Alarm1, Alarm2, Burglary) / P(Burglary, Alarm1) = $0.016/0.072 \sim 0.22$

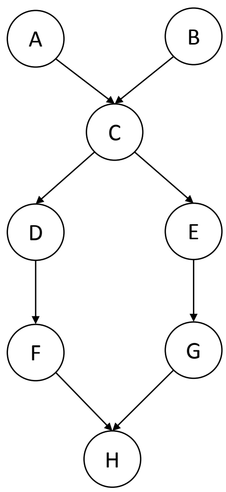
with

P(Alarm1, Alarm2, Burglary) = P(Alarm1) P(Alarm2) P(Burglary | Alarm1, Alarm2) = 0.1 × 0.2 × 0.8 = 0.016

P(Alarm1, ¬ Alarm2, Burglary) = P(Alarm1) P(¬ Alarm2) P(Burglary | Alarm1, ¬ Alarm2) = 0.1 × 0.8 × 0.7 = 0.056

P(Burglary, Alarm1) = P(Alarm1, Alarm2, Burglary) + P(Alarm1, ¬ Alarm2, Burglary) = 0.016 + 0.056 = 0.072

3) Consider the following Bayesian network:



a) Are D and E necessarily independent given evidence about both A and B?

Answer:
No. The path D-C-E is not blocked.

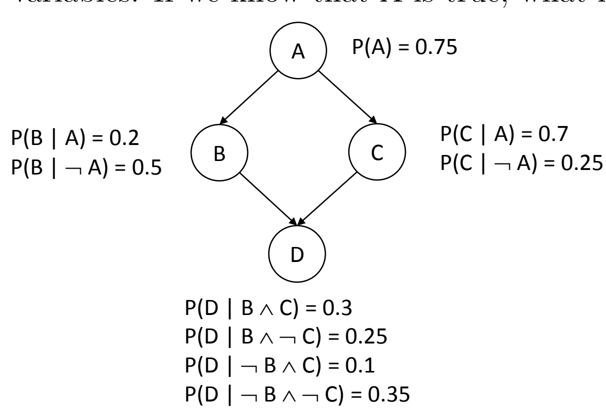
b) Are A and C necessarily independent given evidence about D?

Answer:
No. They are directly dependent. The path A-C is not blocked.

c) Are A and H necessarily independent given evidence about C?

Answer:
Yes. All paths from A to H are blocked.

4) Consider the following Bayesian network. A, B, C, and D are Boolean random variables. If we know that A is true, what is the probability of D being true?



Answer:

$P(D|A)$ = P(A, D) / P(A)

= (P(A, B, C, D) + P(A, B, ¬ C, D) + P(A, ¬ B, C, D) + P(A, ¬ B, ¬ C, D)) / P(A)

= P(B | A) P(C | A) P(D | B, C) + P(B | A) P(¬ C | A) P(D | B, ¬ C) +

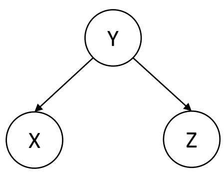
P(¬ B | A) P(C | A) P(D | ¬ B, C) + P(¬ B | A) P(¬ C | A) P(D | ¬ B, ¬ C)

= (0.2 × 0.7 × 0.3) + (0.2 × 0.3 × 0.25) + (0.8 × 0.7 × 0.1) + (0.8 × 0.3 × 0.35)

= 0.042 + 0.015 + 0.056 + 0.084

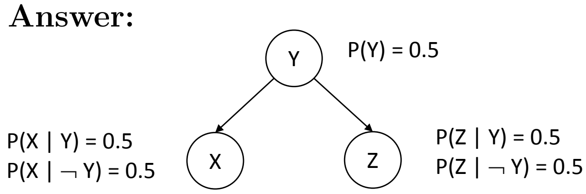
= 0.197

5) For the following Bayesian network



we know that X and Z are not guaranteed to be independent if the value of Y is unknown. This means that, depending on the probabilities, X and Z can be independent or dependent if the value of Y is unknown. Construct probabilities where X and Z are independent if the value of Y is unknown, and show that they are indeed independent.

Answer:



P(X) = P(Y) P(X | Y) + P(¬ Y) P(X | ¬ Y) = 0.5 × 0.5 + 0.5 × 0.5 = 0.5

P(Z) = P(Y) P(Z | Y) + P(¬ Y) P(Z | ¬ Y) = 0.5 × 0.5 + 0.5 × 0.5 = 0.5

P(X, Z) = P(X, Y, Z) + P(X, ¬ Y, Z)

= P(Y) P(X | Y) P(Z | Y) + P(¬ Y) P(X | ¬ Y) P(Z | $\neg Z$)

= 0.5 × 0.5 × 0.5 + 0.5 × 0.5 × 0.5 = 0.25

Therefore, P(X) P(Z) = P(X,Z). We can similarly show that P(X) P(¬ Z) = P(X, ¬ Z), P(¬ X) P(Z) = P(¬ X, Z) and P(¬ X) P(¬ Z) = P(¬ X, ¬ Z) to prove that X and Z are independent if the value of Y is unknown.

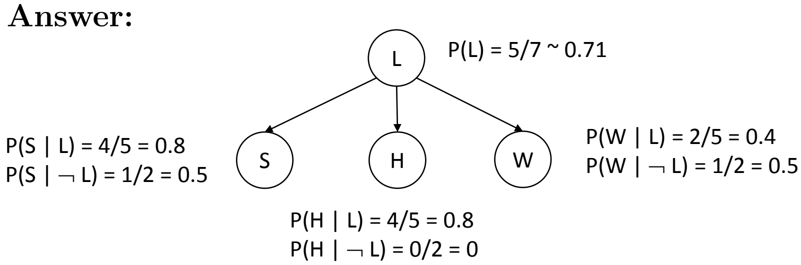
Naive Bayesian Learner

6) A theme park hired you after graduation. Assume that you want to predict when the theme park receives lots of visitors. You gathered the following data:

	Feature 1 Sunny?	Feature 2 High Temperature?	Feature 3 Weekend?	Class Lots of Visitors?
Day 1	yes	yes	yes	yes
Day 2	yes	no	yes	yes
Day 3	no	yes	no	yes
Day 4	yes	yes	no	yes
Day 5	yes	yes	no	yes
Day 6	yes	no	no	no
Day 7	no data since you were on business travel			
Day 8	no	no	yes	no

a) Show the Bayesian network (= hypothesis = model) that a naive Bayesian learner will learn from the data.

Answer:



where S = Sunny, H = High Temperature, W = Weekend and L = Lots of Visitors

b) What's the probability that the learned Bayesian network will predict that the theme park receives lots of visitors on a cloudy and hot weekend day?

Answer:

$P(L \mid \neg S, H, W) = P(L, \neg S, H, W)/P(\neg S, H, W)$

$= 0.045/0.045 = 1$

with

$P(L, \neg S, H, W) = P(L)\ P(\neg S \mid L)\ P(H \mid L)\ P(W \mid L)$

$= 0.71 \times 0.2 \times 0.8 \times 0.4 \sim 0.045$

$P(\neg L, \neg S, H, W) = P(\neg L)\ P(\neg S \mid \neg L)\ P(H \mid \neg L)\ P(W \mid \neg L)$

$= 0.29 \times 0.5 \times 0 \times 0.5 = 0$

$P(\neg S, H, W) = P(L, \neg S, H, W) + P(\neg L, \neg S, H, W) = 0.045 + 0 = 0.045$

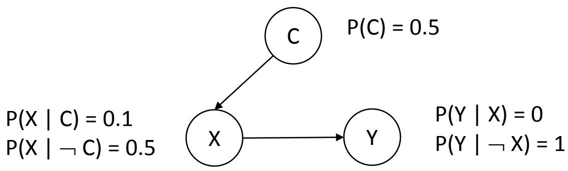
7) Construct an example where a naive Bayesian learner predicts for a feature vector that the predicted class must be true with probability 1 when, in reality, it is false.

Answer:

The previous question gives such an example. It might be the case that there are not many visitors on a cloudy and hot weekend day (even though it does not show up in the data), but the naive Bayesian learner predicts that there will be lots of visitors with probability 1.

8) Give an example of a hypothesis (= model, here: joint probability distribution) that a naive Bayesian learner cannot learn correctly.

Answer:
Consider the Bayesian network



which has the following joint probability distribution:

C	X	Y	
t	t	t	$0.5 \times 0.1 \times 0 = 0$
t	t	f	$0.5 \times 0.1 \times 1 = 0.05$
t	f	t	$0.5 \times 0.9 \times 1 = 0.45$
t	f	f	$0.5 \times 0.9 \times 0 = 0$
f	t	t	$0.5 \times 0.5 \times 0 = 0$
f	t	f	$0.5 \times 0.5 \times 1 = 0.25$
f	f	t	$0.5 \times 0.5 \times 1 = 0.25$
f	f	f	$0.5 \times 0.5 \times 0 = 0$

Lets assume that we are trying predict when C is true, given whether X and Y are true. Given enough random samples drawn from the above joint probability distribution as training data, the naive Bayesian classifier would learn the following:

$\mathrm{P(C)} = 0 + 0.05 + 0.45 + 0 = 0.5$

$\mathrm{P(X \mid C)} = \mathrm{P(X, C)} \;/\; \mathrm{P(C)} = (0 + 0.05)/0.5 = 0.1$

$\mathrm{P(X \mid \neg\, C)} = \mathrm{P(X}, \neg C) \;/\; \mathrm{P(\neg\, C)} = (0 + 0.25)/0.5 = 0.5$

$\mathrm{P(Y \mid C)} = \mathrm{P(Y, C)} \;/\; \mathrm{P(C)} = (0 + 0.45)/0.5 = 0.9$

$\mathrm{P(Y \mid \neg\, C)} = \mathrm{P(Y, \neg\, C)} \;/\; \mathrm{P(\neg\, C)} = (0 + 0.25)/0.5 = 0.5$

which correspond to the following joint probability distribution, different from the original:

C	X	Y	
t	t	t	$0.5 \times 0.1 \times 0.9 = 0.045$
t	t	f	$0.5 \times 0.1 \times 0.1 = 0.005$
t	f	t	$0.5 \times 0.9 \times 0.9 = 0.405$
t	f	f	$0.5 \times 0.9 \times 0.1 = 0.045$
f	t	t	$0.5 \times 0.5 \times 0.5 = 0.125$
f	t	f	$0.5 \times 0.5 \times 0.5 = 0.125$
f	f	t	$0.5 \times 0.5 \times 0.5 = 0.125$
f	f	f	$0.5 \times 0.5 \times 0.5 = 0.125$

9) Give an example where the assumptions that a naive Bayesian learner makes are wrong.

Answer:
The answer to the previous question gives such an example. The naive Bayesian learner assumes that X and Y are independent if the value of C is known when, in reality, they are not.