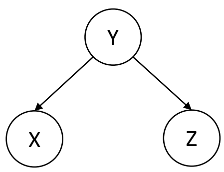
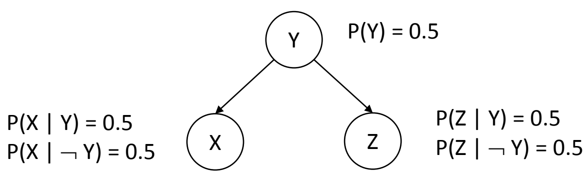
CS360 – Homework #11

Bayesian Networks

1) 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.



$P(X) = P(Y)\ P(X \mid Y) + P(\neg\ Y)\ P(X \mid \neg\ Y) = 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5$

$P(Z) = P(Y)\ P(Z \mid Y) + P(\neg\ Y)\ P(Z \mid \neg\ Y) = 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5$

$P(X, Z) = P(X, Y, Z) + P(X, \neg\ Y, Z)$

$= P(Y)\ P(X \mid Y)\ P(Z \mid Y) + P(\neg\ Y)\ P(X \mid \neg\ Y)\ P(Z \mid \neg Z)$

$= 0.5 \times 0.5 \times 0.5 + 0.5 \times 0.5 \times 0.5 = 0.25$

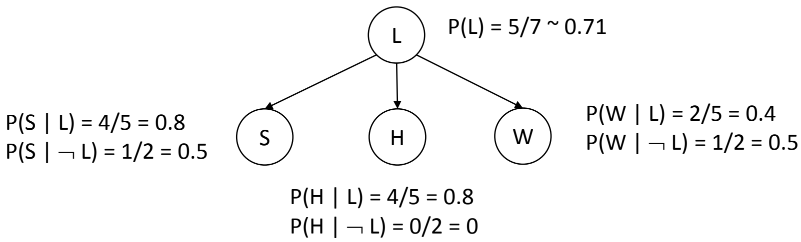
Therefore, P(X) P(Z) = P(X,Z). We can similarly show that P(X) P(¬ Z) = P(X, ¬ Z), P(¬ X) P(Z) = P(¬ X, Z) and P(¬ X) P(¬ Z) = P(¬ X, ¬ Z) to prove that X and Z are independent if the value of Y is unknown.

Naive Bayesian Learner

2) 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.



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?

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

$= 0.045/0.045 = 1$

with

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

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

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

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

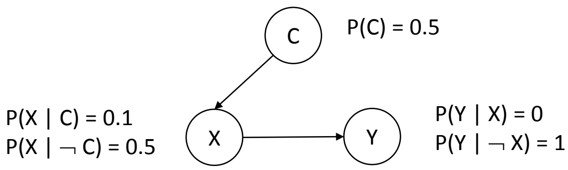
$P(\neg S, H, W) = P(L, \neg S, H, W) + P(\neg L, \neg S, H, W) = 0.045 + 0 = 0.045$

3) 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.

Question 2 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.

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

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:

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

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

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

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

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

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	$0.5 \times 0.5 \times 0.5 = 0.125$

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

The answer to question 4 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.