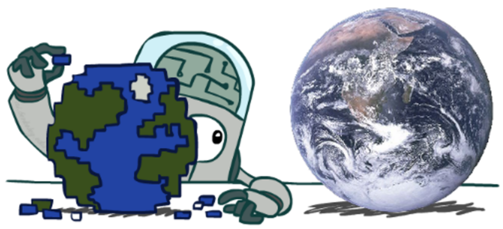
CS440 – Bayesian Networks



Purpose: Models describe how (a portion of) the world works and are always simplifications. We can make simplifications to our probabilistic models to make computations more efficient. Bayesian Networks are a technique to do this.

The purpose of this module is to introduce the concepts of conditional independence and the use of that concept in Bayesian Networks.

Knowledge: This module will help you become familiar with the following content knowledge:

- Independence
- Conditional independence
- Bayesian Networks

Activity1 - Independence and Conditional Independence:
With your group perform the following tasks and answer the questions. You will be reporting your answers back to the class in 30 minutes.

1. Two variables are *independent* of $\forall x, y : P(x, y) = P(x)P(y)$

   This says that their joint distribution factors into two simpler distributions.

   Another form of independence: $\forall x, y : P(x|y) = P(x)$

   Independence of variables X and Y is denoted as $X \perp\!\!\!\perp Y$

   Independence is a simplifying modeling assumption.

   Which pairs of the following variables do you think are independent and which do you think are not independent? {Weather, Traffic, Cavity, Toothache}

   Do you think T and W are independent below? Why or why not?

   $P(T)$

T	P
hot	0.5
cold	0.5

   $P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

   $P(W)$

W	P
sun	0.6
rain	0.4

2. Consider flipping a coin n times and observing whether you get heads or tails each time. Each of these n flips are independent.

   $P(X_1)$

H	0.5
T	0.5

   $P(X_2)$

H	0.5
T	0.5

   $\cdots$

   $P(X_n)$

H	0.5
T	0.5

   What would be the size of the joint distribution for these n variables? Do we need or want this joint distribution table? What saving do we get from independence?

3. There is a less strong version of independence called *conditional independence*. Consider the variables {Cavity, Toothache, Catch} which represent whether someone has a cavity (or not), a toothache (or not) and whether the dental probe catches (possibly signally the existence of a cavity). None of the variables are independent of each other. But if I have a cavity, the probability that the probe catches doesn't depend on whether I have a toothache or not. We will use the notation $+cavity$ to indicate the event that I do have a cavity and $-cavity$ to indicate I do not. Similarly for toothache and catch. So

   $$P(+catch| + toothache, +cavity) = P(+catch| + cavity)$$

   The same independence holds if I don't have a cavity:

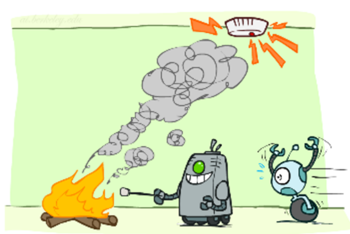
   $$P(+catch| + toothache, -cavity) = P(+catch| - cavity)$$

   This makes Catch conditionally independent of Toothache given Cavity:

   $$P(Catch|Toothache, Cavity) = P(Catch|Cavity)$$

   Consider the variables {Traffic, Umbrella, Raining} indicating whether or not there is (or is not) traffic, whether I have (or not) an umbrella, and whether it is (or is not) raining. Are any of these variable independent? What two could be conditionally independent given the third?

   Consider the variables {Fire, Smoke, Alarm}.

   Are any of these variables independent? What two could be conditionally independent given the third?

4. Just like independence, conditional independence can simplify our model. Remember the chain rule:

$$P(X_1, X_2, ..., X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)...$$

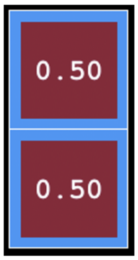
So with the chain rule,

P(Traffic,Rain,Umbrella) = P(Rain) P(Traffic| Rain) P(Umbrella |Traffic,Rain)

How can this be simplified with the assumption that Umbrella and Traffic are conditionally independent given Rain?

Activity2 - Bayesian Networks:
With your group perform the following tasks and answer the questions. You will be reporting your answers back to the class in 30 minutes.

1. Consider a very simple version of looking for a ghost using noisy sensors where there are only two cells:

   Let the variable T indicate that the sensor is red in the top cell.
   Let the variable B indicate that the sensor is red in the bottom cell.
   Let the variable G indicate that the ghost is in the top cell.

   Givens:

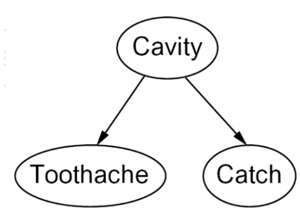
   $P(+g) = 0.5$
   $P(-g) = 0.5$
   $P(+t \mid +g) = 0.8$
   $P(+t \mid -g) = 0.4$
   $P(+b \mid +g) = 0.4$
   $P(+b \mid -g) = 0.8$

   The chain rule tells us that: $P(T,B,G) = P(G)\ P(T \mid G)\ P(B \mid G,T)$

   Do we have enough information to create the joint distribution here? If not, what information are we missing?

   Conditional independence will come to the rescue. Each sensor depends only on where the ghost is located. That means, the two sensors are conditionally independent, given the ghost position. Use this new information to generate the joint distribution P(T,B,G).

2. We can use a graphical notation to illustrate the conditional independence relationships. The nodes of the graph will represent variables. The arcs in the graph will represent interactions between the variables. For now, you can imagine that the arrows means causation (although in general they don't).



What is this graph telling us about conditional independence?

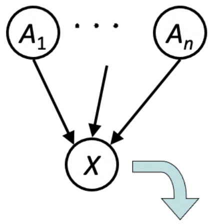
Draw a causal graph with the following variables:

{Traffic, Rain, LowPressure, RoofDrips, Ballgame, Cavity}

Draw a causal graph with the following variables:

{Burglary, AlarmGoesOff, MaryCalls, JohnCalls, Earthquake!}
(presumably your neighbors Mary and John would call you if they hear your alarm)

3. A Bayesian Network is a set of nodes, one per variable X. It is a directed, acyclic graph and each node contains a conditional probability table for all possible values of its parents.

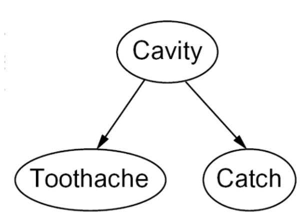


$$P(X|A_1 \ldots A_n)$$

The Bayesian network implicitly encodes the joint distribution since

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

Look at this example:



The table P(Cavity) resides in the Cavity node. The table P(ToothAche | Cavity) resides in the Toothache node. The table P(Catch | Cavity) resides in the Catch node. How would you use the BN to compute P(+cavity, -toothache, +catch)?

4. Look at the BN for coin flips. Since the variables are all absolutely independent we can represent this with a BN with no arcs.

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

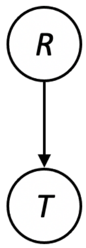
...

$P(X_n)$

h	0.5
t	0.5

Use the BN to compute $P(h, h, t, h)$

Here is the BN for Rain and Traffic:



$P(R)$

+r	1/4
-r	3/4

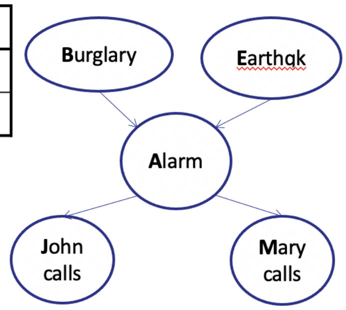
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

Use the BN to compute $P(+r, -t)$

Here is the BN for the alarm:

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

Use the BN to compute the probability that there is both a burglary and earthquake, no alarm but John and Mary both call anyway.