

CS 2810 April 19 Dec 4

Admin:

"high five" warmup

You need a spreadsheet applicatoin today

TRACE participation update

HW9 due Weds April 27

Bayes Nets!

- compute multiple target vars from multiple evidence vars
 - $P(ABC|XYZ) = P(ABCXYZ) / P(XYZ)$
- conditional independence
- bayes net notation
- computing conditional probabilities
 - via spreadsheet ("computer" method)
 - algebraically
 - (15 mins of next lesson)

(enjoy Bayes Nets?

see "Probabilistic Graphical Models" Daphne Koller & Coursera course)

Conditional Independence

(algebraic) definition:

We say that X, Y conditionally independent (given Z) if:

$$P(X|Y, Z) = P(X|Z) \text{ and } P(Y|X, Z) = P(Y|Z)$$

Example: F and T are conditionally independent given W

- Marathon (F) or casted weather (day before)
- Observed (W) eather day of marathon
- Average (T) ime of runners on course

If the forecasted weather is "good" then run times will be lower.

- in general, F and T are dependent

Given that we observe the actual weather, then the forecast no longer informs average run time.

- after observing the particular W , F and T are independent



(intuitive) definition: the only way X and Y influence each other is through Z

Bayesian Network (Bayes Net)



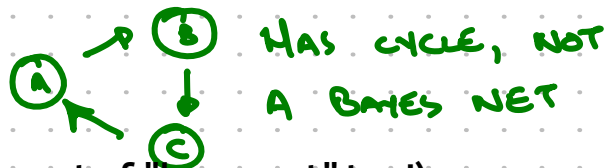
(formally):

A directed, **acyclic** graph which represents conditional distributions / independences between a set of random variables.

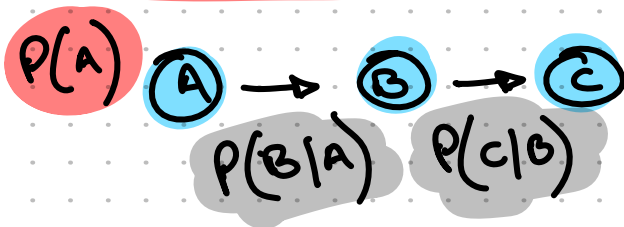
each node represents a random variable

directed edges represent conditional distributions

any node without inward edges has prob specified (its part of "Bayes net" too!)



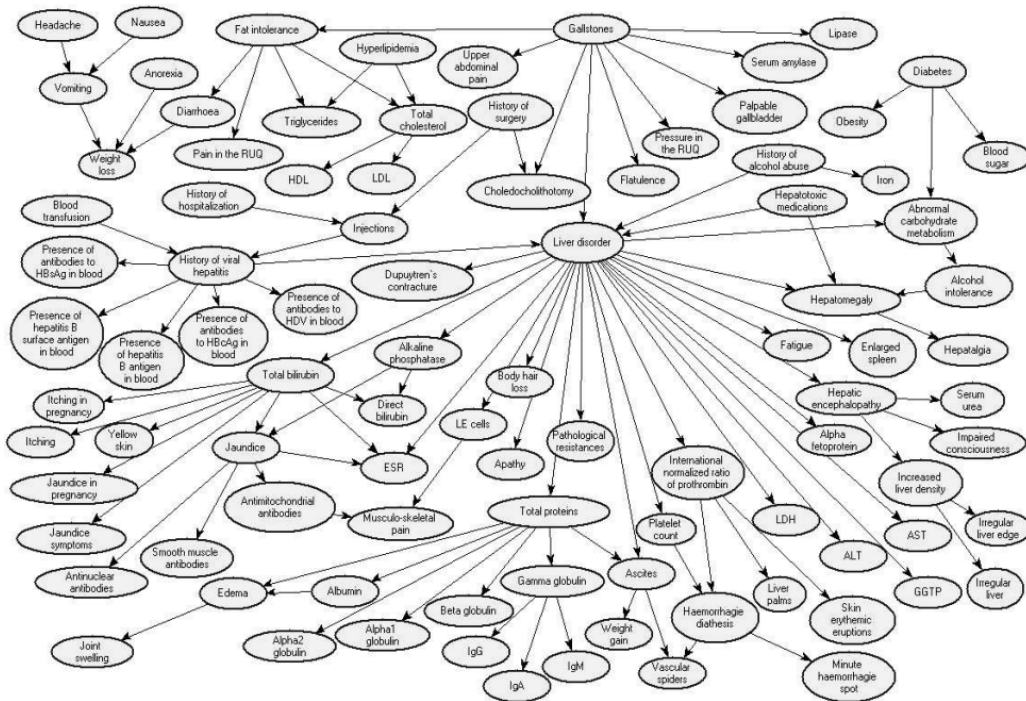
HAS CYCLE, NOT
A BAYES NET



(informally):

a network which describes how random variables influence each other. can be used to compute conditional probabilities of interest

WHAT ARE BAYES NETS
GOOD FOR?



Bayes nets allow us to incorporate multiple pieces of evidence into some conditional prob of interest:

given a person has:

- symptom 4
- symptom 11
- risk factor 7

whats the prob of liver disorder?

source: <https://sites.pitt.edu/~druzdzel/psfiles/cbmi99a.pdf>

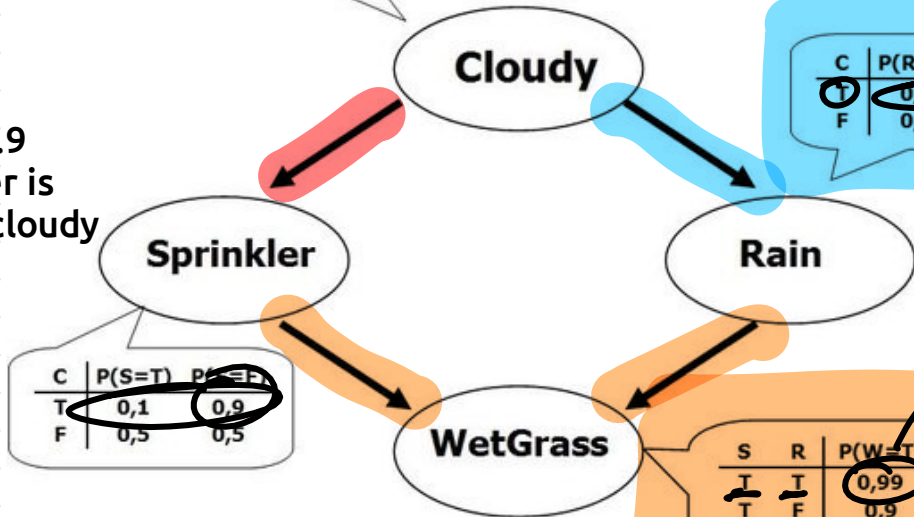
ANATOMY OF BAYES NET

Prob cloudy = True is 50%

	P(C=T)	P(C=F)
	0,5	0,5

$P(R=T|C=T) = .8$

$P(S=F|C=T) = .9$
 Prob sprinkler is off given it's cloudy out is 90%



	C	P(R=T)	P(R=F)
T		0,8	0,2
F		0,2	0,8

$P(W=T|S=T, R=T) = .99$
 Prob that grass is wet given sprinkler is on and its raining is 99%

	C	P(S=T)	P(S=F)
T		0,1	0,9
F		0,5	0,5

	S	R	P(W=T)	P(W=F)
T	T		0,99	0,01
T	F		0,9	0,1
F	T		0,9	0,1
F	F		0,0	1,0

BAYES NET NOTATION (OUR CONVENTION)

Each random variable is denoted with a capital letter (T for Thief). Each outcome in sample space has its own lowercase letter :

t0 = no thief
t1 = thief

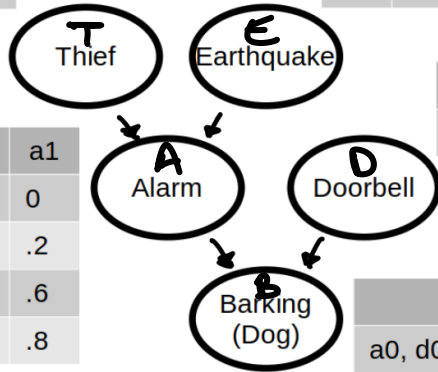
t0	t1
.99	.01

e0	e1
.95	.05

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

d0	d1
.8	.2

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99



(quick) ICA 1:

what's prob of earthquake?

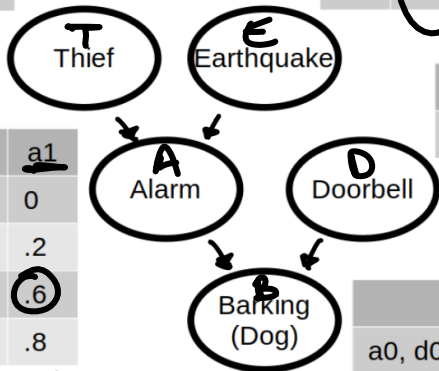
given a thief in house, but no earthquake, what's prob alarm goes off?

interpretation question:
- is alarm better at detecting thieves or earthquakes?

- which sound bothers the dog more, the alarm or doorbell?

t0	t1
.99	.01

e0	e1
.95	.05



d0	d1
.8	.2

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99

$$P(A|TE)$$

(quick) ICA 1:

what's prob of earthquake?

$$P(e_1) = .05$$

given a thief in house, but no earthquake, what's prob alarm goes off?

$$P(a_1 | t_1, e_0) = .6$$

interpretation question:

- is alarm better at detecting thieves or earthquakes?

- which sound bothers the dog more, the alarm or doorbell?

$$P(a_1 | t_1, e_0) > P(a_1 | t_0, e_1)$$

$$P(b_1 | a_0, d_1) > P(b_1 | a_1, d_0)$$

In Class Assignment 2:

Estimate / intuite the four probabilities below, which are greater / lesser / equal to other probs?

What is the prob of thief?

$$P(t) = .01$$

Given that alarm is going off, what is prob of thief?

$$\rightarrow P(t|a) > P(t)$$

Given that alarm is going off & dog is barking, what is prob of thief?

$$P(t|a) = P(t|a, b)$$

Given that alarm is going off, dog is barking & earthquake, what is prob of thief?

$$P(t|a, b, e) < P(t|a, b)$$

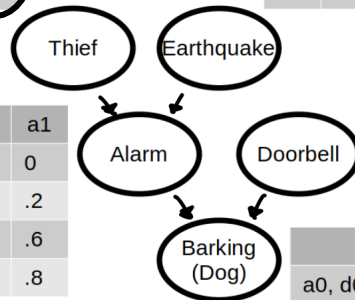
t0	a1
.99	.01

e0	e1
.95	.05

d0	d1
.8	.2

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99



How do we compute conditional probabilities from a Bayes Net?

With a computer:

Step 1: Rewrite conditional probability without conditional ✓

Step 2(c): In a spreadsheet, compute prob of every possible combination of outputs for all vars

Step 3(c): Compute the needed probabilities from step 1 via marginalization

With algebra:

Step 1: Rewrite conditional probability without conditional

Step 2(a): rewrite each conditional probability using only probabilities given in Bayes Net

- add variables via marginalization

$$P(A) = \sum_b P(A, b)$$

- factor joint distributions into given conditional probabilities:

$$P(A, B) = P(B|A) P(A)$$

- utilize given independence relationships between variables

$$P(A, B) = P(A) P(B)$$

Step 3(a): plug in values

NEXT
LESSON

Step 1: write conditional probabilities as ratio of (not conditional) probabilities

$$P(A|B) = \frac{P(AB)}{P(B)}$$

$$P(\text{ABC} | \text{XYZ}) = \frac{P(\text{ABC XYZ})}{P(\text{XYZ})}$$

↑ TARGET VARS ↑ EVIDENCE VARS

→ PROB OF TARGET AND EVIDENCE TOGETHER

→ PROB OF EVIDENCE

(ex: Given that alarm is going off & dog is barking, what is prob of thief?)

$$P(t_i | a_i b_i) = \frac{P(t_i a_i b_i)}{P(a_i b_i)}$$

$$P(A|B)P(B) = P(AB)$$

With a computer ...

Step 2(c): In a spreadsheet, compute prob of every possible combination of outputs for all vars

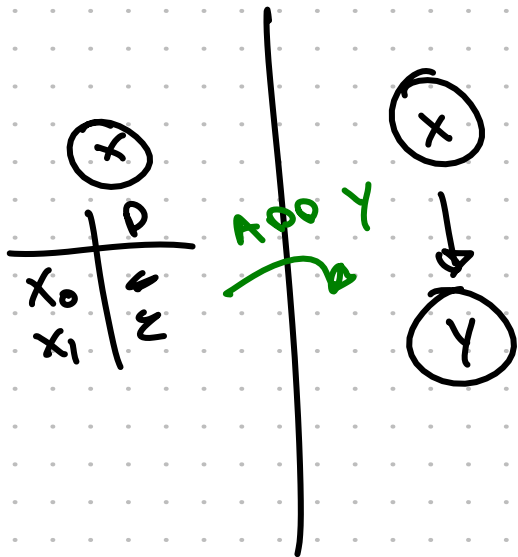
("JOINT DISTRIBUTION TABLE")

B: Barking	D: Doorbell	A: Alarm	T: Thief	E: Earthquake	P(BDATE)
b0	d0	a0	t0	e0	0.7524
b0	d0	a0	t0	e1	0.03168
b0	d0	a0	t1	e0	0.00304
b0	d0	a0	t1	e1	8E-05
b0	d0	a1	t0	e0	0
b0	d0	a1	t0	e1	0.00396
b0	d0	a1	t1	e0	0.00228
b0	d0	a1	t1	e1	0.00016
b0	d1	a0	t0	e0	0.03762
b0	d1	a0	t0	e1	0.001584
b0	d1	a0	t1	e0	0.000152
b0	d1	a0	t1	e1	4E-06
b0	d1	a1	t0	e0	0
b0	d1	a1	t0	e1	1.98E-05
b0	d1	a1	t1	e0	1.14E-05
b0	d1	a1	t1	e1	8E-07
b1	d0	a0	t0	e0	0
b1	d0	a0	t0	e1	0

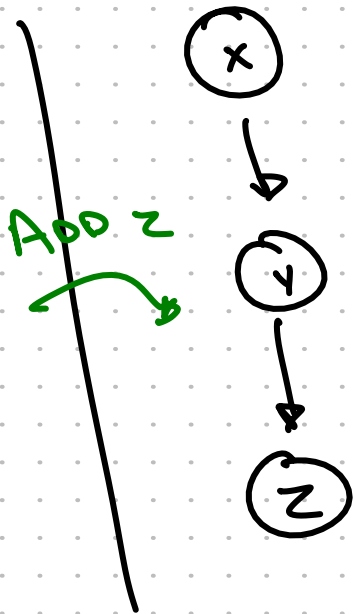
EVERY
COMBINATION
VARS

PROB

PRODUCING A JOINT TABLE ITERATIVELY



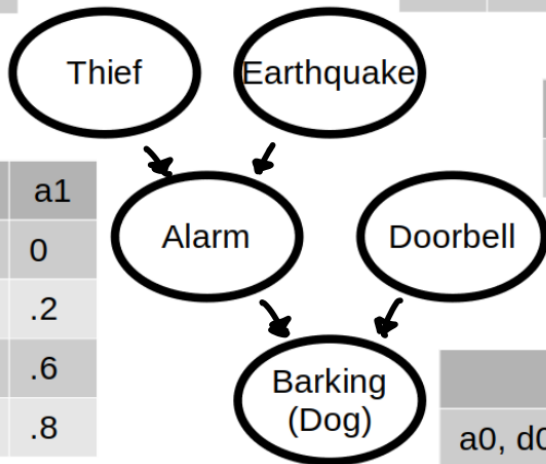
X_0	Y_0	D
X_0	Y_0	D
X_1	Y_0	D
X_1	Y_1	D



	D_{100}
X_0, Y_0, Z_0	D
\vdots	D
X_1, Y_1, Z_1	D

t0	t1
.99	.01

e0	e1
.95	.05



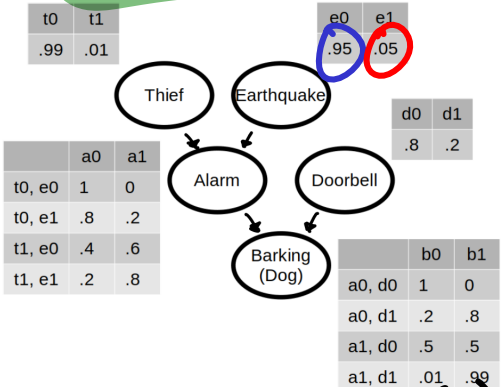
d0	d1
.8	.2

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99

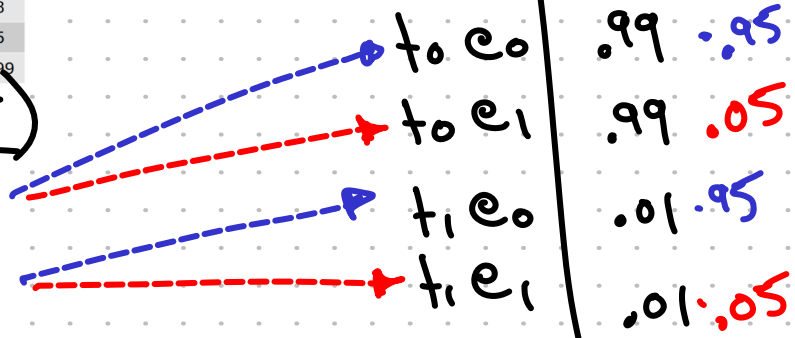
PRODUCING JOINT TABLE (ADDING INDEPENDENT NODES)

SINCE T, E ARE INDEPENDENT
 $P(TE) = P(T)P(E)$

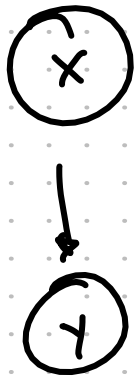


T	P(T)
t0	.99
t1	.01

T	E	P(TE)
t0	e0	.99 .95
t0	e1	.99 .05
t1	e0	.01 .95
t1	e1	.01 .05



PRODUCING JOINT TABLE (DEPENDENT NODES)



	Y_0	Y_1
X_0	$1/3$	$2/3$
X_1	$1/4$	$3/4$

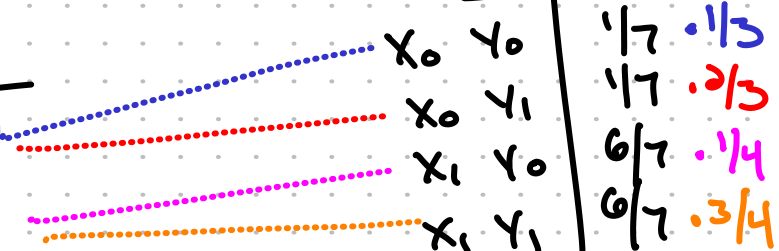
THESE ARE
 $P(Y|X)$
 VALUES

SINCE Y DEPENDS ON X
 $P(X, Y) = P(Y|X) P(X)$

X	P_{MARG}
X_0	$1/7$
X_1	$6/7$

X	Y	$P(X, Y)$
X_0	Y_0	$1/7 \cdot 1/3$
X_0	Y_1	$1/7 \cdot 2/3$
X_1	Y_0	$6/7 \cdot 1/4$
X_1	Y_1	$6/7 \cdot 3/4$

$$P(X_0, Y_0) = P(Y_0|X_0) P(X_0)$$



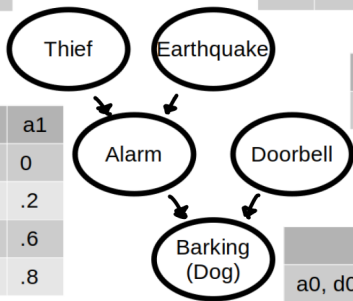
t0	t1
.99	.01

e0	e1
.95	.05

d0	d1
.8	.2

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99



In Class Exercise (don't submit):

Build the joint distribution table for the bayes net on the left.

(You needn't submit for credit. You can check your work with the given final answer csv on website)

$$P(T, E) = P(T | E) P(E) = P(T) P(E)$$

$$P(A, T, E) = P(A | T, E) P(T, E)$$

$$P(D | A, T, E) = P(D | A, T, E) P(A, T, E) \quad (D \text{ is independent})$$

$$P(B | D, A, T, E) = P(B | D, A, T, E) P(D, A, T, E) \quad (B \text{ is conditional independent TE given DA})$$

X	Y	Z	PROB
x_0	y_0	z_0	$\frac{1}{4}$
x_0	y_0	z_1	0
x_0	y_1	z_0	0
x_0	y_1	z_1	$\frac{1}{8}$
x_1	y_0	z_0	$\frac{3}{8}$
x_1	y_0	z_1	0
x_1	y_1	z_0	0
x_1	y_1	z_1	$\frac{1}{4}$

MARGINALIZING IN JOINT TABLE (step 3c)

COMPUTE $P(x_0 z_0)$

$$= \sum_Y P(x_0 Y z_0)$$

$$= P(x_0 y_0 z_0) + P(x_0 y_1 z_0)$$

$$= \frac{1}{4} + 0 = \frac{1}{4}$$

X	Y	Z	Pras
x_0	y_0	z_0	$1/4$
x_0	y_0	z_1	0
x_0	y_1	z_0	0
x_0	y_1	z_1	$1/8$
x_1	y_0	z_0	$3/8$
x_1	y_0	z_1	0
x_1	y_1	z_0	0
x_1	y_1	z_1	$1/4$

MARGINALIZING IN JOINT TABLE (step 3c)

QUICK PRACTICE

COMPUTE $P(y_1, x_1)$

$$= P(x_1, y_1, z_0) + P(x_1, y_1, z_1) = 1/4$$

COMPUTE $P(x_0)$

$$= P(x_0, y_0, z_0) + P(x_0, y_0, z_1) + P(x_0, y_1, z_0) + P(x_0, y_1, z_1)$$

Putting it all together:

Step 1: Rewrite conditional probability without conditional

Step 2(c): In a spreadsheet, compute prob of every possible combination of outputs for all vars

Step 3(c): Compute the needed probabilities from step 1 via marginalization

Example:

Given alarm is going off and dog is barking, what is the probability of a thief?

$$\textcircled{3} \quad P(t_1 | a_1, b_1) = \frac{P(t_1, a_1, b_1)}{P(a_1, b_1)} = .0036478 / .009568 = .381$$

In Class Exercise 3:

Explicitly compute each of the following

1. What is the prob of thief? $P(t_1) = .01$

2. Given that alarm is going off, what is prob of thief?

$$P(t_1 | a_1) = \frac{P(t_1, a_1)}{P(a_1)} = .381$$

3. Given that alarm is going off & dog is barking, what is prob of thief?

4. Given that alarm is going off, dog is barking & earthquake, what is prob of thief?

.381 (SEE PREV)

Answer each question below with one sentence (please avoid algebraic motivations and appeal to our intuition):

- Why is the prob of 2 greater than the prob of 1?
- Why is the prob of 3 equal to the prob of 2?
- Why is the prob of 4 less than the prob of 2?

$$P(t_1 | a_1, b_1, e_1) = \frac{P(t_1, a_1, b_1, e_1)}{P(a_1, b_1, e_1)} = ?$$

How do we compute conditional probabilities from a Bayes Net?

With a computer:

Step 1: Rewrite conditional probability without conditional

Step 2(c): In a spreadsheet, compute prob of every possible combination of outputs for all vars

Step 3(c): Compute the needed probabilities from step 1 via marginalization

With algebra:

Step 1: Rewrite conditional probability without conditional

Step 2(a): rewrite each conditional probability using only probabilities given in Bayes Net

- add variables via marginalization

$$P(A) = \sum_b P(A, b)$$

- factor joint distributions into given conditional probabilities:

$$P(A, B) = P(B|A) P(A)$$

- utilize given independence relationships between variables

$$P(A, B) = P(A) P(B)$$

Step 3(a): plug in values

NEXT
LESSON

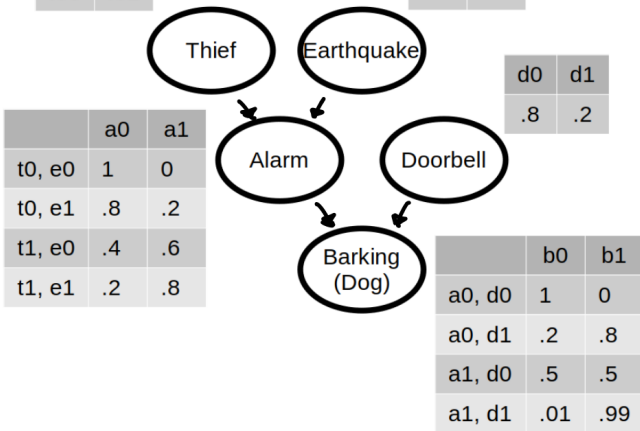
COMPUTE

$P(a_0)$

(PROB ALARM IS OFF)

t0	t1
.99	.01

e0	e1
.95	.05



d0	d1
.8	.2

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99

$$P(a_0) = \sum_{t,e} P(a_0 | t, e)$$

$$= P(a_0 | t_0, e_0) P(t_0) P(e_0)$$

$$+ P(a_0 | t_0, e_1) P(t_0) P(e_1)$$

$$+ P(a_0 | t_1, e_0) P(t_1) P(e_0)$$

$$+ P(a_0 | t_1, e_1) P(t_1) P(e_1)$$

$$= 1 \cdot (.99) \cdot (.95) + .8 \cdot (.99) \cdot (.05) + .4 \cdot (.01) \cdot (.95) + .2 \cdot (.01) \cdot (.05) = .984$$

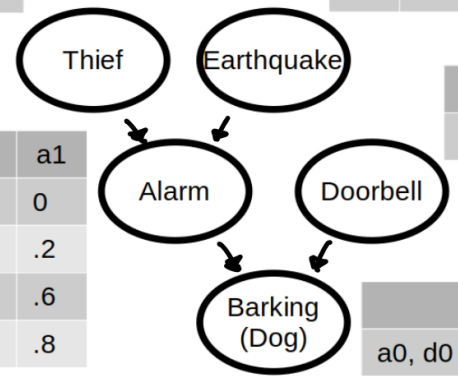
t0	t1
.99	.01

e0	e1
.95	.05

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

d0	d1
.8	.2

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99



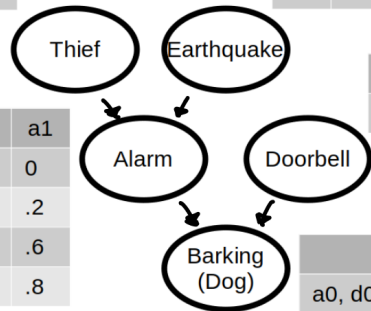
COMPUTE

$P(b_0)$

(PROB DOG NOT BARKING)

t0	t1
.99	.01

e0	e1
.95	.05



d0	d1
.8	.2

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99

Skip

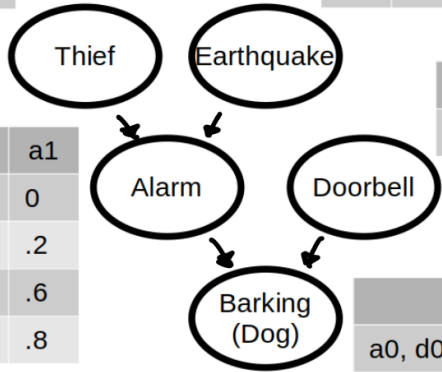
$$\begin{aligned}
 P(b_0) &= \sum_{ad} P(b_0 | a d) \\
 &= P(b_0 | a_0 d_0) P(a_0) P(d_0) \\
 &+ P(b_0 | a_0 d_1) P(a_0) P(d_1) \\
 &+ P(b_0 | a_1 d_0) P(a_1) P(d_0) \\
 &+ P(b_0 | a_1 d_1) P(a_1) P(d_1)
 \end{aligned}$$

t0	t1
.99	.01

e0	e1
.95	.05

	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

d0	d1
.8	.2



	b0	b1
a0, d0	1	0
a0, d1	.2	.8
a1, d0	.5	.5
a1, d1	.01	.99

$$\begin{aligned}
 &= 1 \cdot (.984) \cdot (.8) \quad (.2) \cdot (.984) \cdot (.2) \\
 &+ (.5) \cdot (.016) \cdot (.8) \quad (.01) \cdot (.016) \cdot (.2) = .833
 \end{aligned}$$

EXTRA: NOT ON HW OR QUIZ

TOPOLOGICAL SORT OF DIRECTED GRAPH

ORDER NODES SO THAT IF EDGE x, y EXISTS THEN x IS IN LIST BEFORE y



WE MUST ADD NODES
IN BAYES NET IN
TOPO ORDERING