CS 2810 April 19 Dec 4

Admin:
"high five"
TRACE participation update ~ 50% participation
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 "Probabilisitic 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)orecasted 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 and dependent

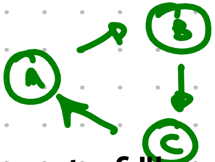
Given that we observe the actual weather, then the forecast no longer informs average run time.
- after observing the particular W, F and T are independent

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

Bayesian Network (Bayes Net)

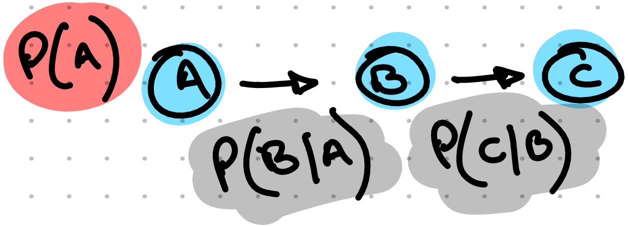
$$A \rightarrow B \rightarrow C$$

(formally):
A directed, acyclic graph which represents conditional distributions / independences between a set of random variables.



HAS CYCLE, NOT A BAYES NET

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!)

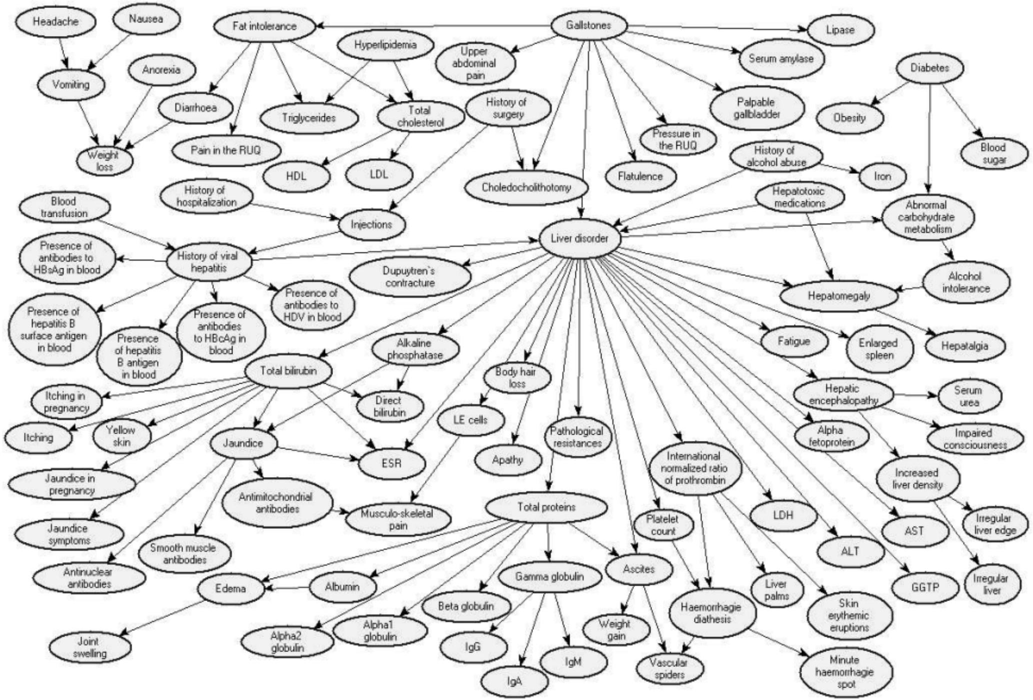


$$P(AB) = P(B|A)\,P(A)$$

$$P(ABC) = P(C|AB)\,P(AB)$$

(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?



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

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

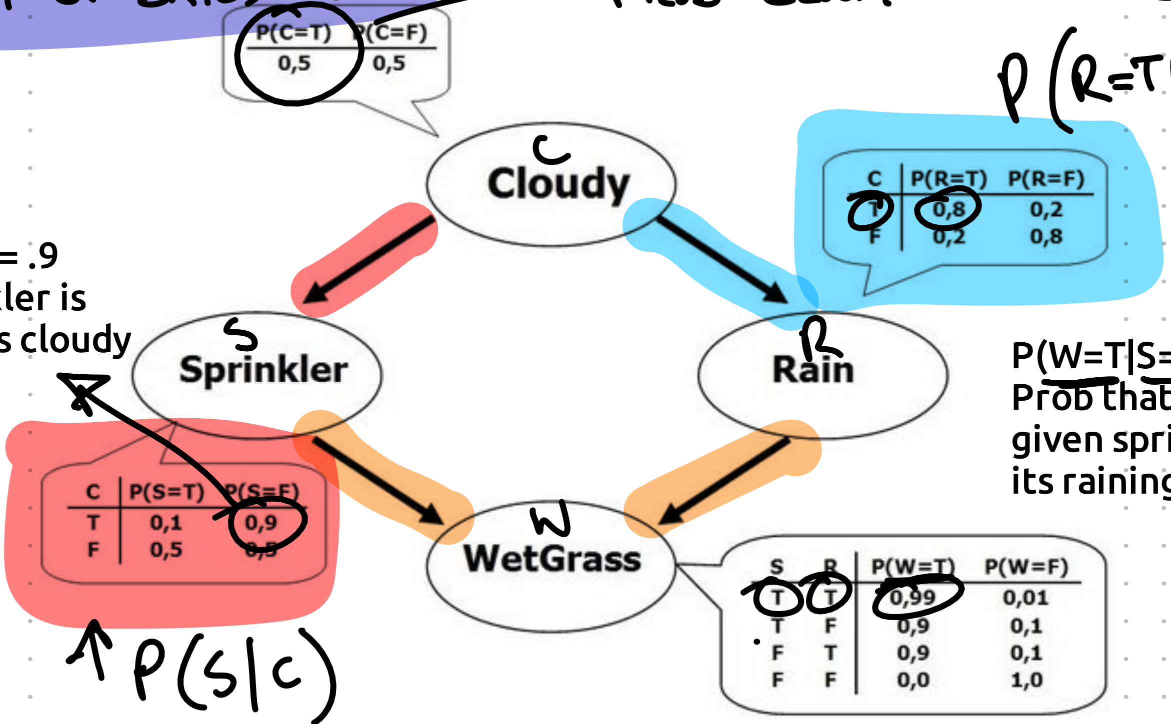
given a person has:
- symptom 4
- symptom 11
- risk factor 7

whats the prob of liver disorder?

Anatomy of Bayes Net

Prob Cloudy = True is 50%

P(R=T|C=T)



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

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

source: https://towardsdatascience.com/introduction-to-bayesian-networks-81031eeed94e

Bayes Net Notation (our convention)

t0 → event no thief

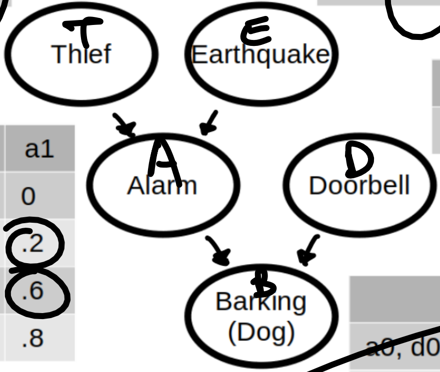
t1 → event thief

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



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

(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?

$.2 = p(a_1 | t_0 e_1) < p(a_1 | t_1 e_0) = .6$

$.8 = p(b_1 | a_0 d_1) > p(b_1 | a_1 d_0) = .5$

In Class Assignment 2:
Estimate / intuite the four probabilities below, which are greater / lesser / equal to other probs?

1. What is the prob of thief? $P(t1) = .01$

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

$P(t1|a1) > P(t1)$. intuition: a1 and t1 positive correlated

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

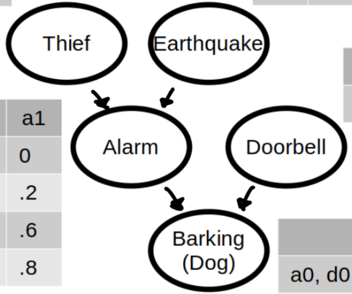
$P(t1|a1, b1) = p(t1|a1)$

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

$P(t1|a1, b1, e1) < P(t1| a1, b1)$

t0	t1
.99	.01

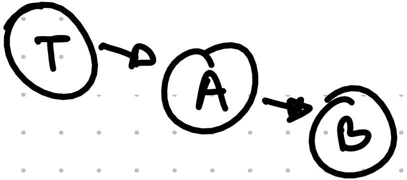
e0	e1
.95	.05



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



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): Computer 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

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

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

PROB OF TARGET AND EVIDENCE TOGETHER

PROB OF EVIDENCE

TARGET VARS

EVIDENCE VARS

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

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

With a computer …

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

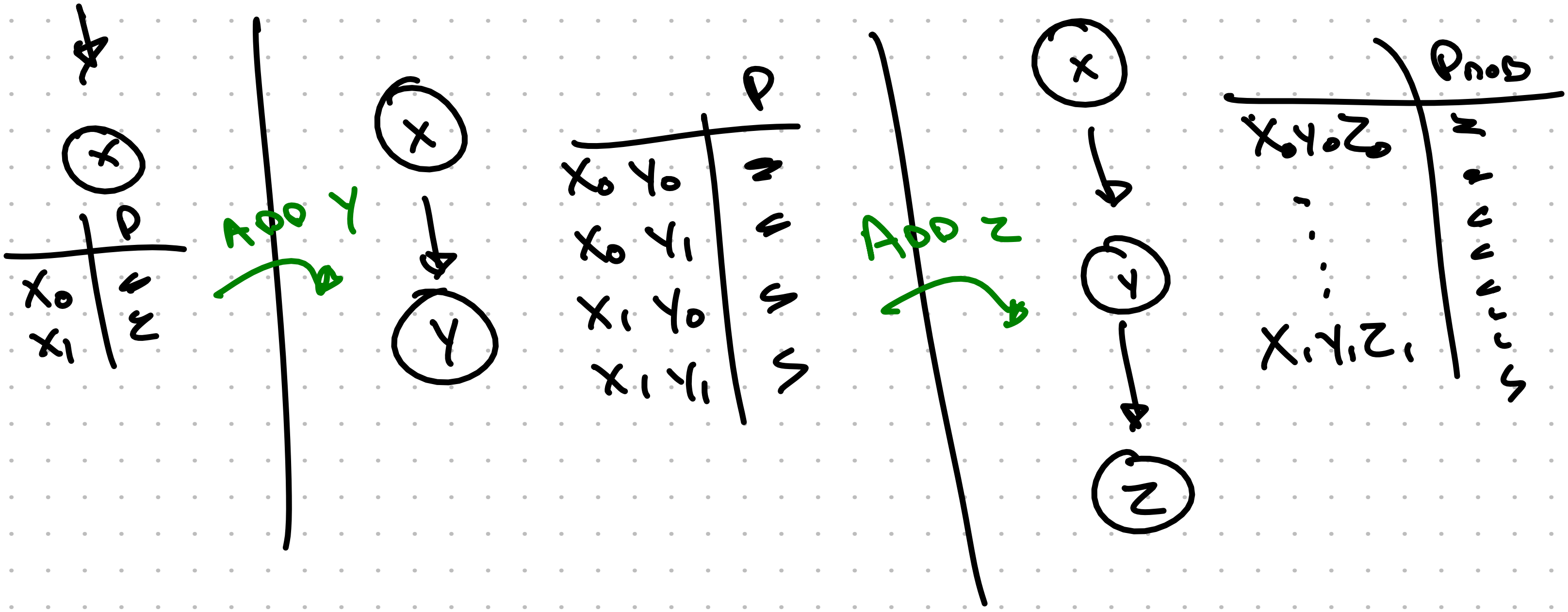
("Joint Distribution Table")

Every combination vars

B: Barking	D: Doorbell	A: Alarm	T: Thief	E: Earthqua	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

Prob

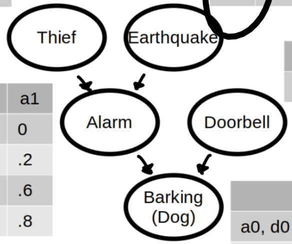
Producing a Joint Table Iteratively



Producing Joint Table (Adding Independent Nodes)

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

Since T, E are independent

$P(TE) = P(T)P(E)$

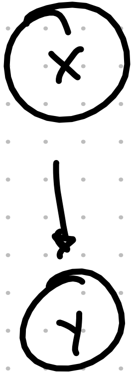
T	P(T)
t_0	.99
t_1	.01

T E	P(TE)
$t_0 e_0$	.99 .95
$t_0 e_1$	.99 .05
$t_1 e_0$	.01 .95
$t_1 e_1$	.01 · .05

$\leftarrow P(t_0 e_0)$

$= P(t_0) P(e_0)$

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

$P(Y_0|X_0) = 1/3$

Since Y depends on X

$P(X\,Y) = P(Y|X)\,P(X)$

X	Prob
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	$6/7 \cdot 3/4$

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

$P(T, E) = P(T|E)\ P(E) = P(T)\ P(E)$ since E, T indept

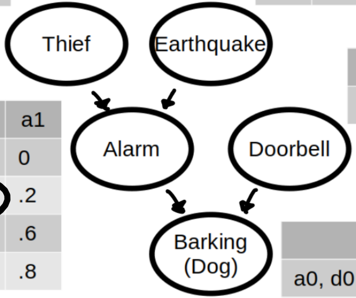
In Class Exercise (don't submit):

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)

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

$P(A|TE)$

$P(ATE) = P(A|TE)\ P(TE) = P(A|TE)\ P(T)P(E)$ since T, E are independent

$P(DATE) = P(D|ATE)\ P(ATE) = P(D)\ P(ATE)$ since D is independent of ATE

$P(BDATE) = P(B|DATE)\ P(DATE) = P(B|DA)P(DATE)$ since B is conditional indep of TE given DA

Marginalizing in Joint Table
(step 3c)

X	Y	Z	Prob
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

Compute $P(x_0 z_0)$

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

$= 1/4 + 0 = 1/4$

MARGINALIZING IN JOINT TABLE
(step 3c)

X	Y	Z	PROB
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

QUICK PRACTICE

COMPUTE $P(y_1, x_1) = 1/4$

COMPUTE $P(x_0) = 1/4 + 1/8 = 3/8$

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?

$$p(t_1 \mid a_1, b_1) = \frac{p(t_1, a_1, b_1)}{p(a_1, b_1)} \approx \frac{.0036}{.00957} \approx .381$$

p(t1, a1, b1)	0.0036478
p(a1, b1)	0.009568
p(t1|a1, b1)	0.38125

In Class Exercise 3:

$P(a1, t1) = 0.0061$
$p(a1) = 0.016$
$p(t1|a1) = 0.38125$

Explicitly compute each of the following

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

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

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

3. Given that alarm is going off & dog is barking, what is prob of thief? $P(t_1 | a_1 b_1) = .381$

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

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?

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

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

P(t1, a1, e1, b1)=	0.0002392
P(a1, e1, b1) =	0.0061594
P(t1|a1, e1, b1) =	0.038834951456311

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): Computer 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)

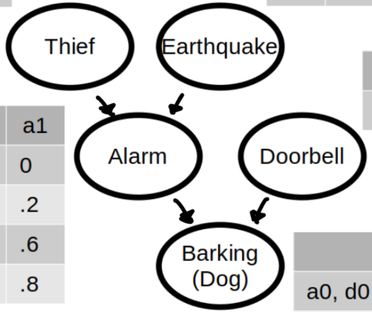
Step 3(a): plug in values

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

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

$$= P(a_0 \mid t_0 e_0)\, P(t_0)\, P(e_0)$$

$$P(a_0 \mid t_0 e_1)\, P(t_0)\, P(e_1)$$

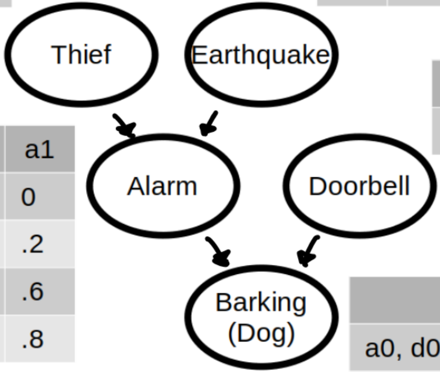
$$P(a_0 \mid t_1 e_0)\, P(t_1)\, P(e_0)$$

$$P(a_0 \mid t_1 e_1)\, P(t_1)\, P(e_1)$$

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

t0	t1
.99	.01

e0	e1
.95	.05



d0	d1
.8	.2

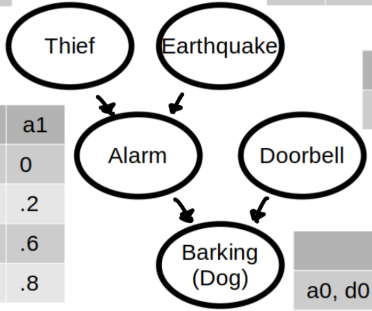
	a0	a1
t0, e0	1	0
t0, e1	.8	.2
t1, e0	.4	.6
t1, e1	.2	.8

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

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

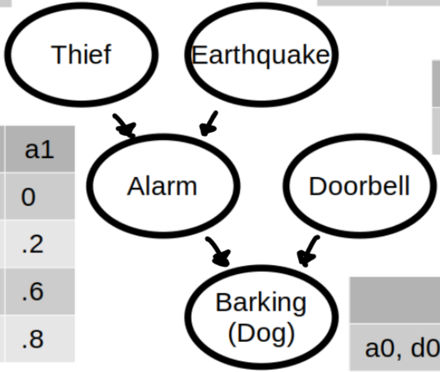
$$P(b_0) = \sum_{ad} P(b_0\, a\, d)$$

$$= P(b_0|a_0 d_0)\,P(a_0)\,P(d_0)$$
$$+ P(b_0|a_0 d_1)\,P(a_0)\,P(d_1)$$
$$+ P(b_0|a_1 d_0)\,P(a_1)\,P(d_0)$$
$$+ P(b_0|a_1 d_1)\,P(a_1)\,P(d_1)$$

$$= 1\,(.984)\,(.8) \quad (.2)\,(.984)\,(.2)$$
$$(.5)\,(.016)\,(.8) \quad (.01)\,(.016)\,(.2) = .833$$

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

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

A,B,C IS TOPO SORTED

A,C,B IS NOT TOPO SORTED

WE MUST ADD NODES IN BAYES NET IN TOPO ORDERING