CSC242: Introduction to Artificial Intelligence

Lecture 4.5

Announcements

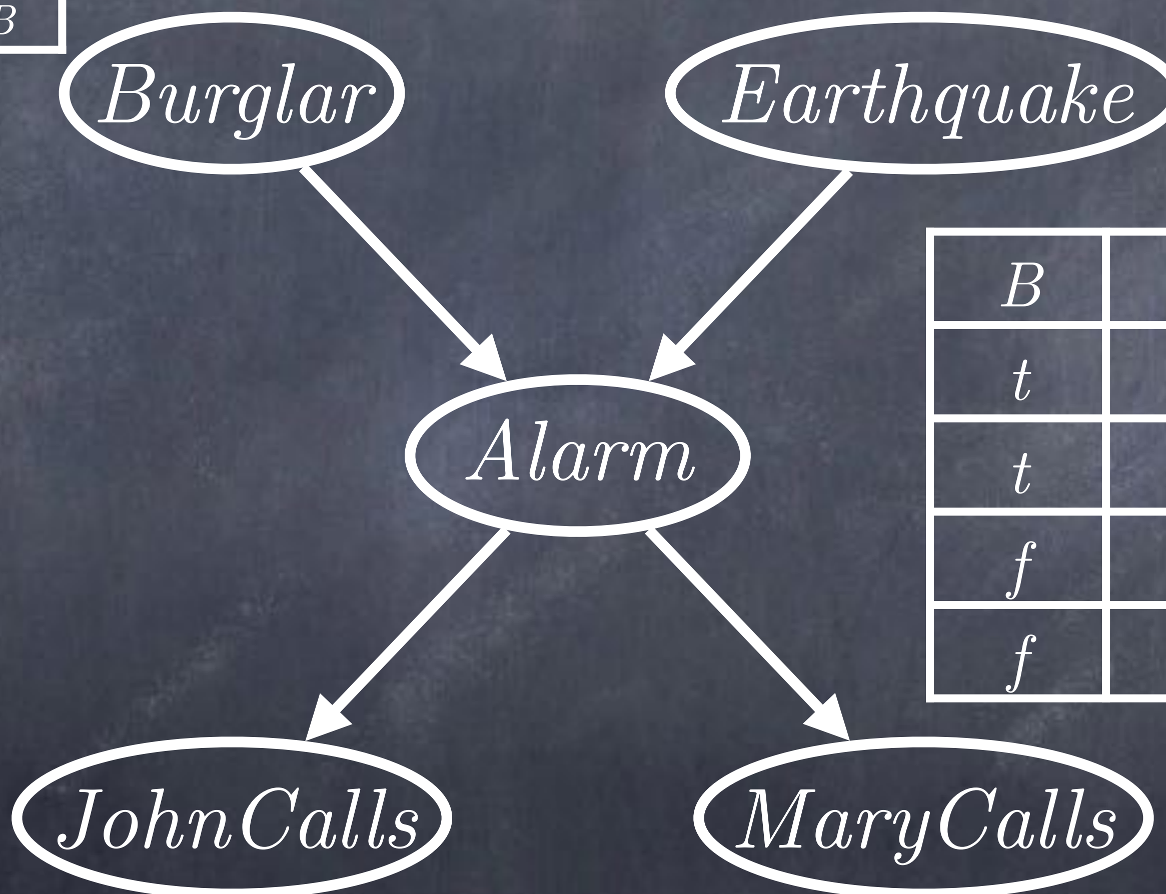
- Unit 4 Exam: Thu 27 Apr 940AM
- Project 4 due Thu 27 Apr 1159PM
- Final Exam: Thu 11 May 1600
  Douglass Ballroom (not Dewey 1-101)
  - **BRING ID**

	Bags	Agent, process, disease, ...
	Candies	Actions, effects, symptoms, results of tests, ...
$D_1 =$ $D_2 =$ $D_3 =$	Observations	
Goal	Predict next candy	Predict agent's next move Predict next output of process Predict disease given symptoms and tests

Parameter Learning (in Bayesian Networks)

$P(B)$
Θ_B

$P(E)$
Θ_E



B	E	$P(A)$
t	t	Θ_{A1}
t	f	Θ_{A2}
f	t	Θ_{A3}
f	f	Θ_{A4}

A	$P(J)$
t	Θ_{J1}
f	Θ_{J2}

A	$P(E)$
t	Θ_{E1}
f	Θ_{E2}

Independent Identically Distributed (i.i.d.)

- Probability of a sample is independent of any previous samples

$$\mathbf{P}(D_i | D_{i-1}, D_{i-2}, \ldots) = \mathbf{P}(D_i)$$

- Probability distribution doesn't change among samples

$$\mathbf{P}(D_i) = \mathbf{P}(D_{i-1}) = \mathbf{P}(D_{i-2}) = \cdots$$

Maximum Likelihood Hypothesis

$$\operatorname*{argmax}_{\Theta} P(\mathbf{d} \mid h_{\Theta})$$

h_Θ

$P(F{=}cherry)$	$P(F{=}lime)$
Θ	1-Θ

$Flavor$

Maximum Likelihood Hypothesis

$$L(\mathbf{d} \mid h_\Theta) = c \log \Theta + l \log(1 - \Theta)$$

$$\operatorname*{argmax}_{\Theta} L(\mathbf{d} \mid h_\Theta) = \frac{c}{c + l} = \frac{c}{N}$$

$h_{\Theta,\Theta_1,\Theta_2}$

$P(F{=}c)$	$P(F{=}l)$
Θ	1-Θ

F	$P(W{=}r \mid F)$	$P(W{=}g \mid F)$
c	Θ_1	1-Θ_1
l	Θ_2	1-Θ_2

Maximum Likelihood Hypothesis

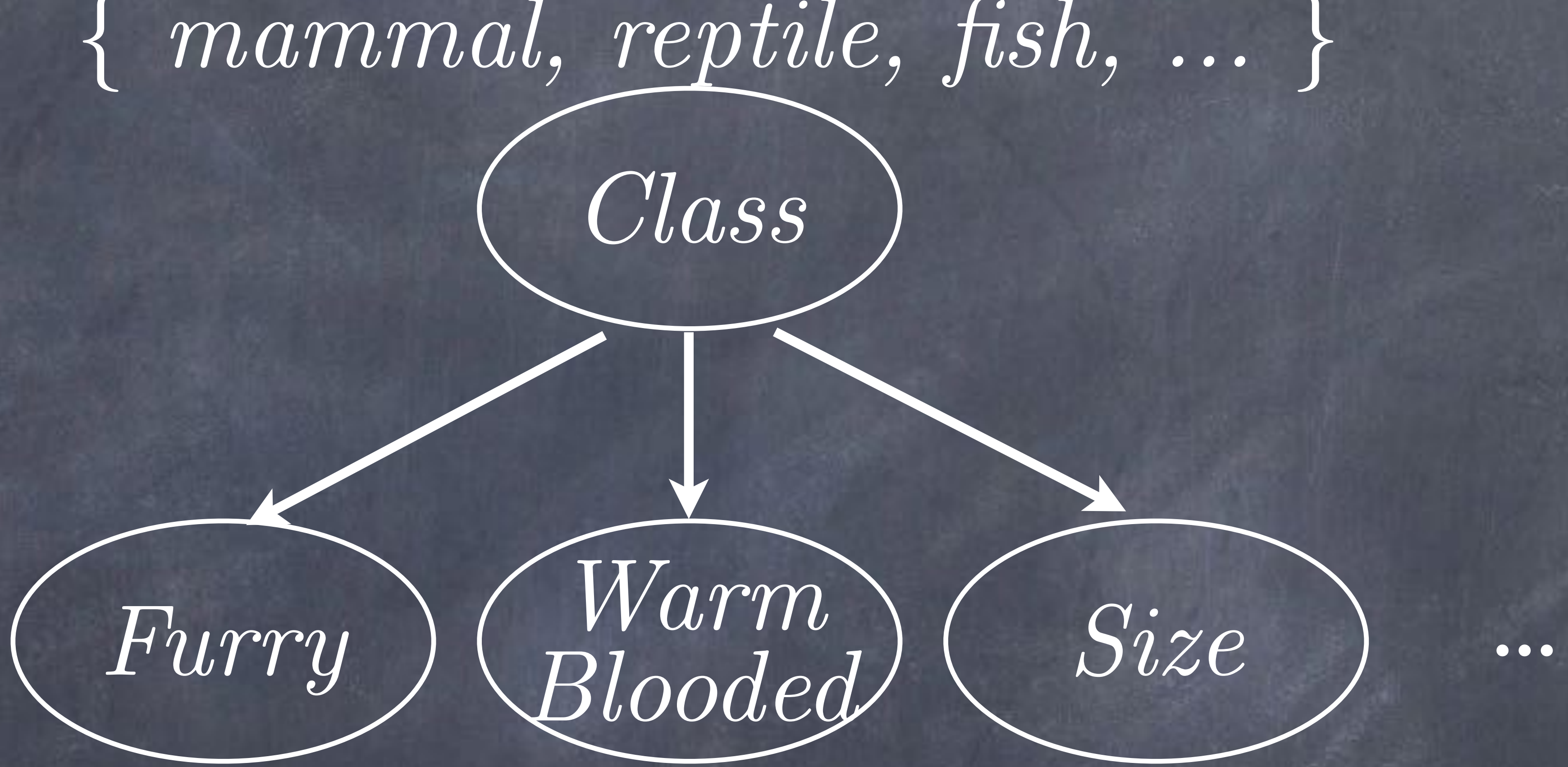
$$\Theta = \frac{c}{c+l} = \frac{c}{N}$$

$$\Theta_1 = \frac{r_c}{r_c + g_c} = \frac{r_c}{c}$$

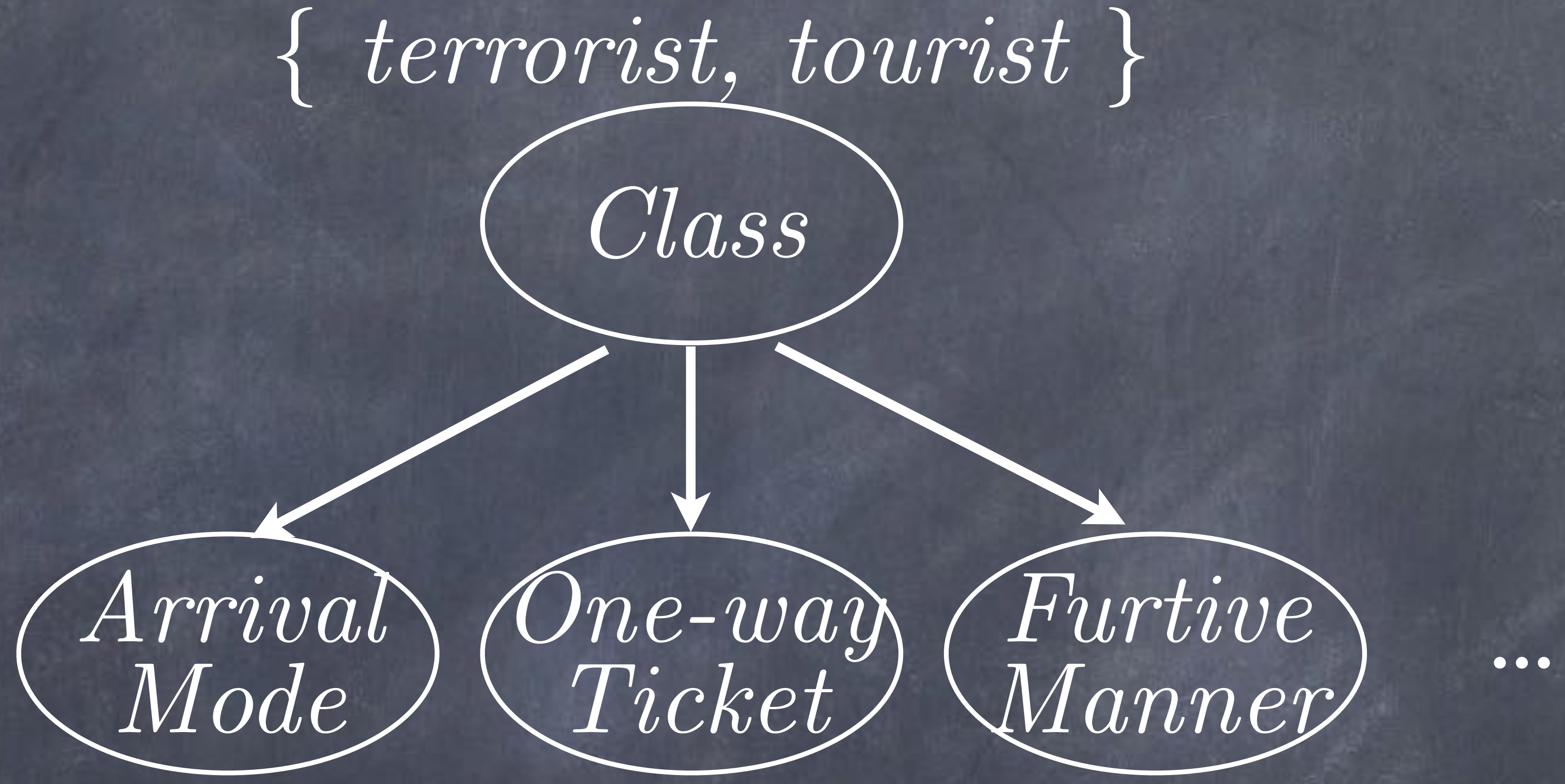
$$\Theta_2 = \frac{r_l}{r_l + g_l} = \frac{r_l}{l}$$

Observed frequencies are the BEST hypothesis, in terms of maximizing the likelihood of the data.

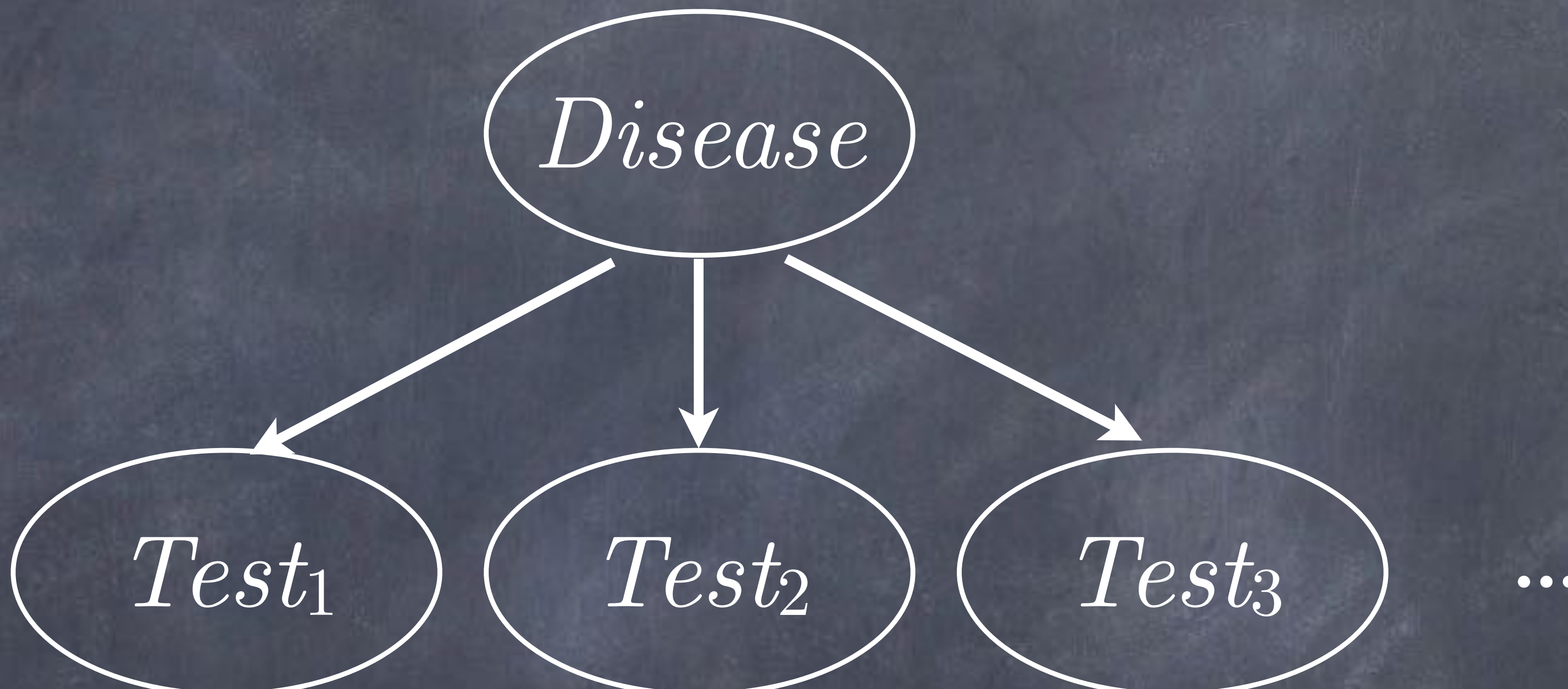
Naive Bayes Models



Naive Bayes Models



Naive Bayes Models



Learning Naive Bayes Models

- Naive Bayes model with n Boolean attributes requires $2n+1$ parameters
- Maximum likelihood hypothesis can be found with no search
  - Probabilities are observed frequencies
  - Scales to large problems
- Robust to noisy or missing data

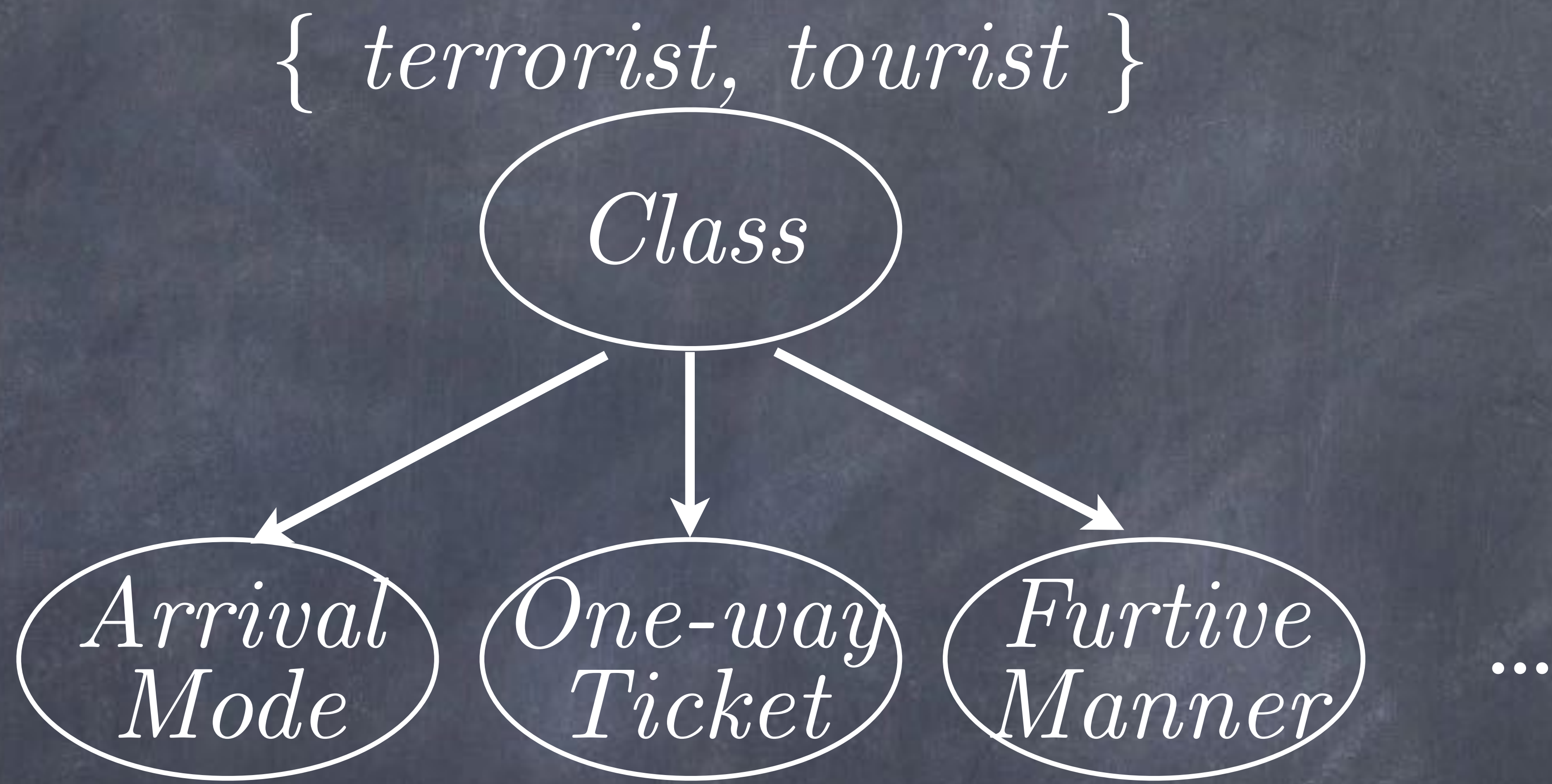
Naive Bayes Classifier

New input: x_1, ..., x_n

$$\begin{aligned}\mathbf{P}(C \mid x_1, \ldots, x_n) &= \alpha\, \mathbf{P}(x_1, \ldots, x_n \mid C)\, \mathbf{P}(C) \\ &= \alpha\, \mathbf{P}(C) \prod_i \mathbf{P}(x_i | C)\end{aligned}$$

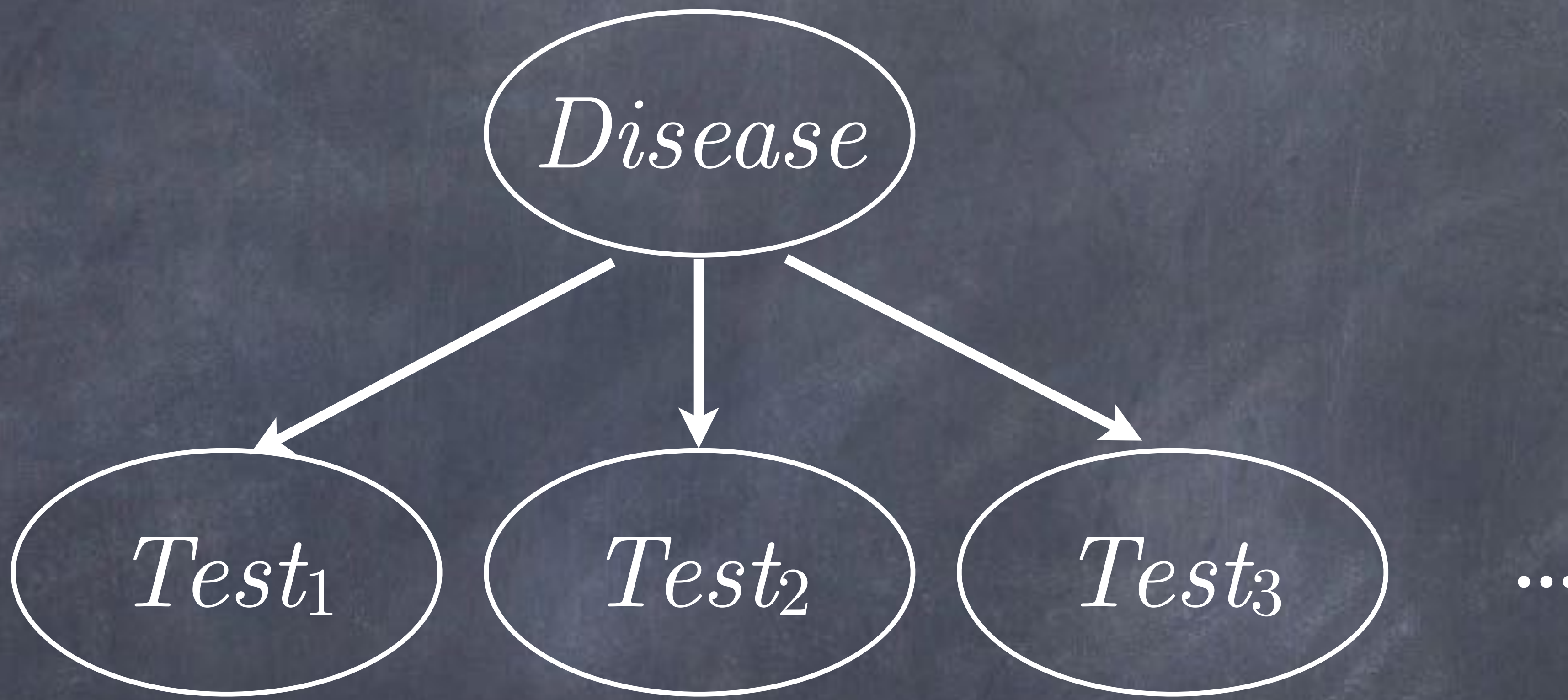
Parameter Learning in Bayesian Networks

- Can learn the CPTs for a Bayes Net from observations that include values for all variables
- Finding maximum likelihood parameters decomposes into separate problems, one for each parameter
- Parameter values for a variable given its parents are the observed frequencies

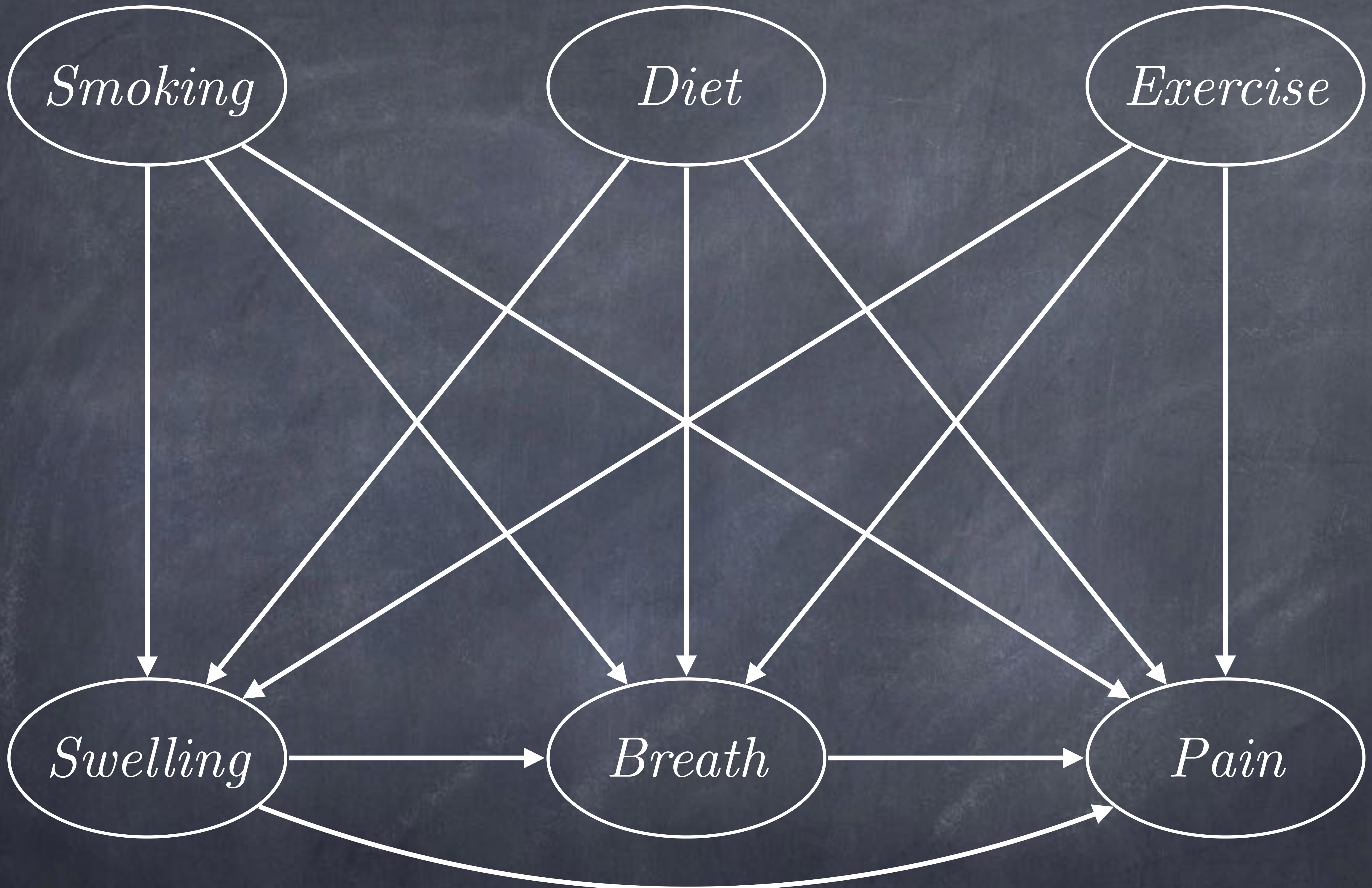

{ terrorist, tourist }
Class
Arrival Mode
One-way Ticket
Furtive Manner
...

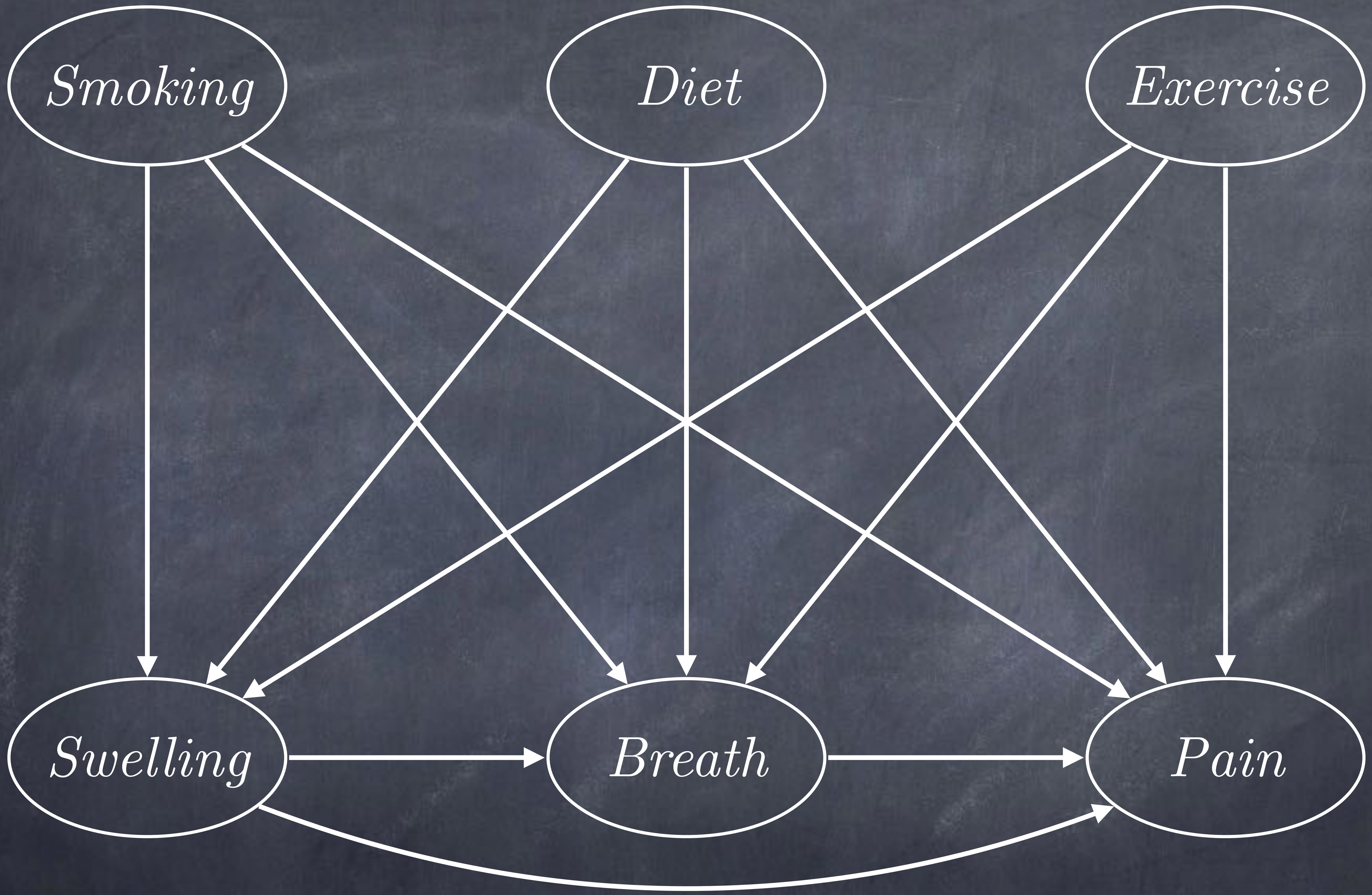
Arrival	*One-Way*	*Furtive*	...	*Class*
taxi	*yes*	*very*	...	*terrorist*
car	*no*	*none*	...	*tourist*
car	*yes*	*very*	...	*terrorist*
car	*yes*	*some*	...	*tourist*
walk	*yes*	*none*	...	*student*
bus	*no*	*some*	...	*tourist*

Arrival	*One-Way*	*Furtive*	...	*Class*
taxi	*yes*	*very*	...	*terrorist?*
car	*no*	*none*	...	*tourist?*
car	*yes*	*very*	...	*terrorist?*
car	*yes*	*some*	...	*tourist?*
walk	*yes*	*none*	...	*student?*
bus	*no*	*some*	...	*tourist?*

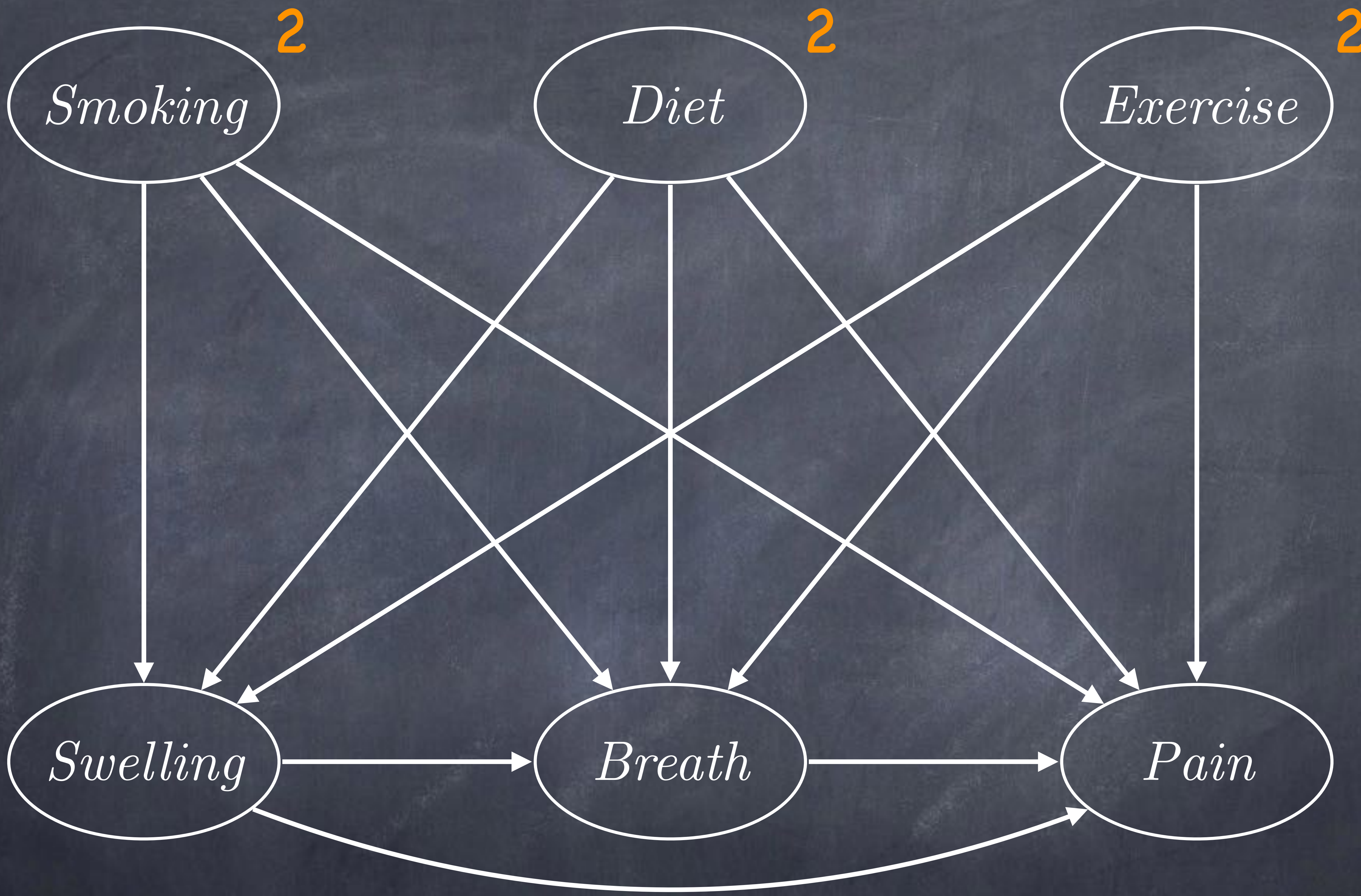

Disease
Test1
Test2
Test3
...

$Test$	$Test2$	$Test3$	$\ldots$	$Disease$
T	F	T	$\ldots$	$?$
T	F	F	$\ldots$	$?$
F	F	T	$\ldots$	$?$
T	T	T	$\ldots$	$?$
F	T	F	$\ldots$	$?$
T	F	T	$\ldots$	$?$

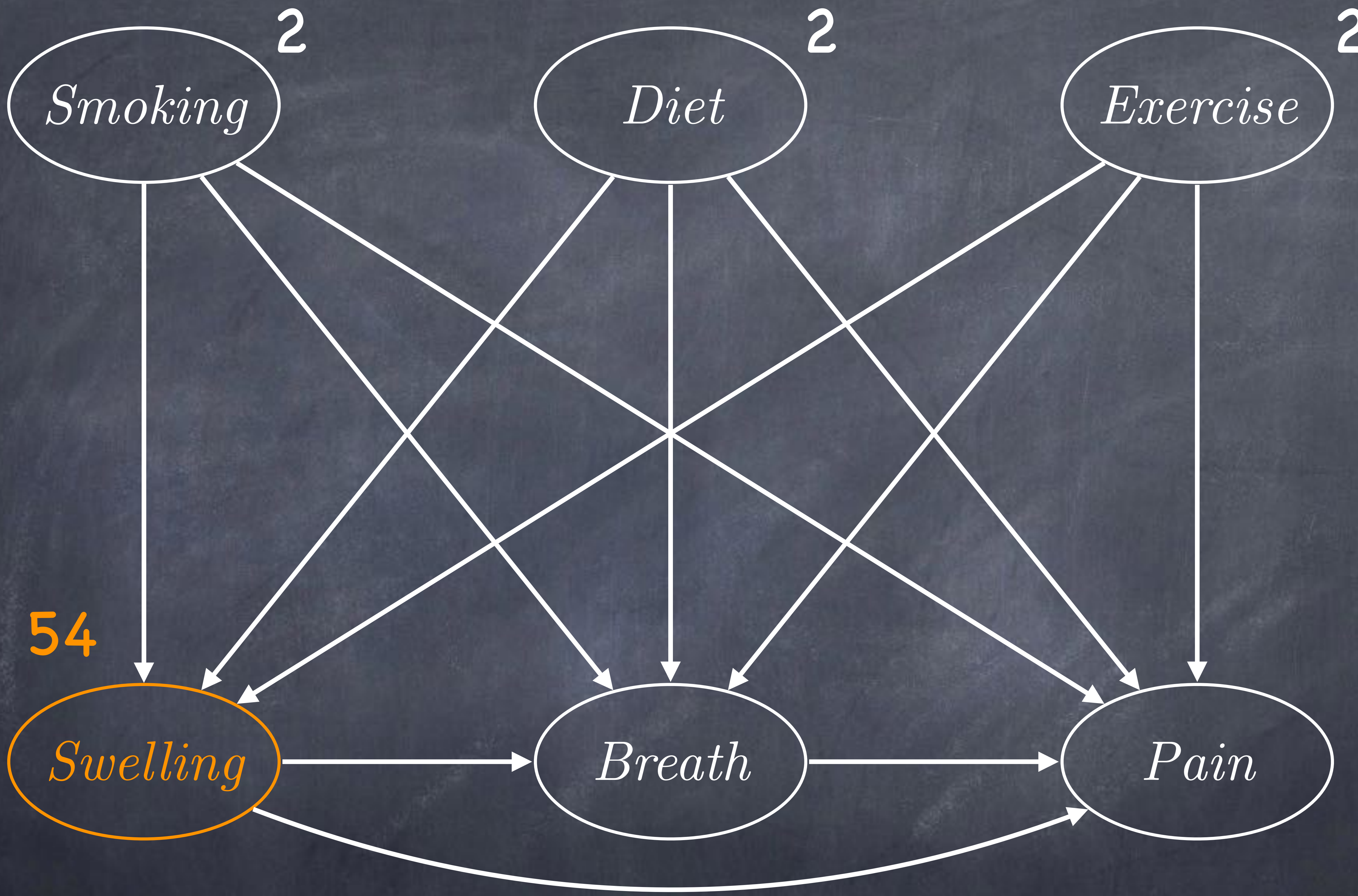

Smoking
Diet
Exercise
Swelling
Breath
Pain



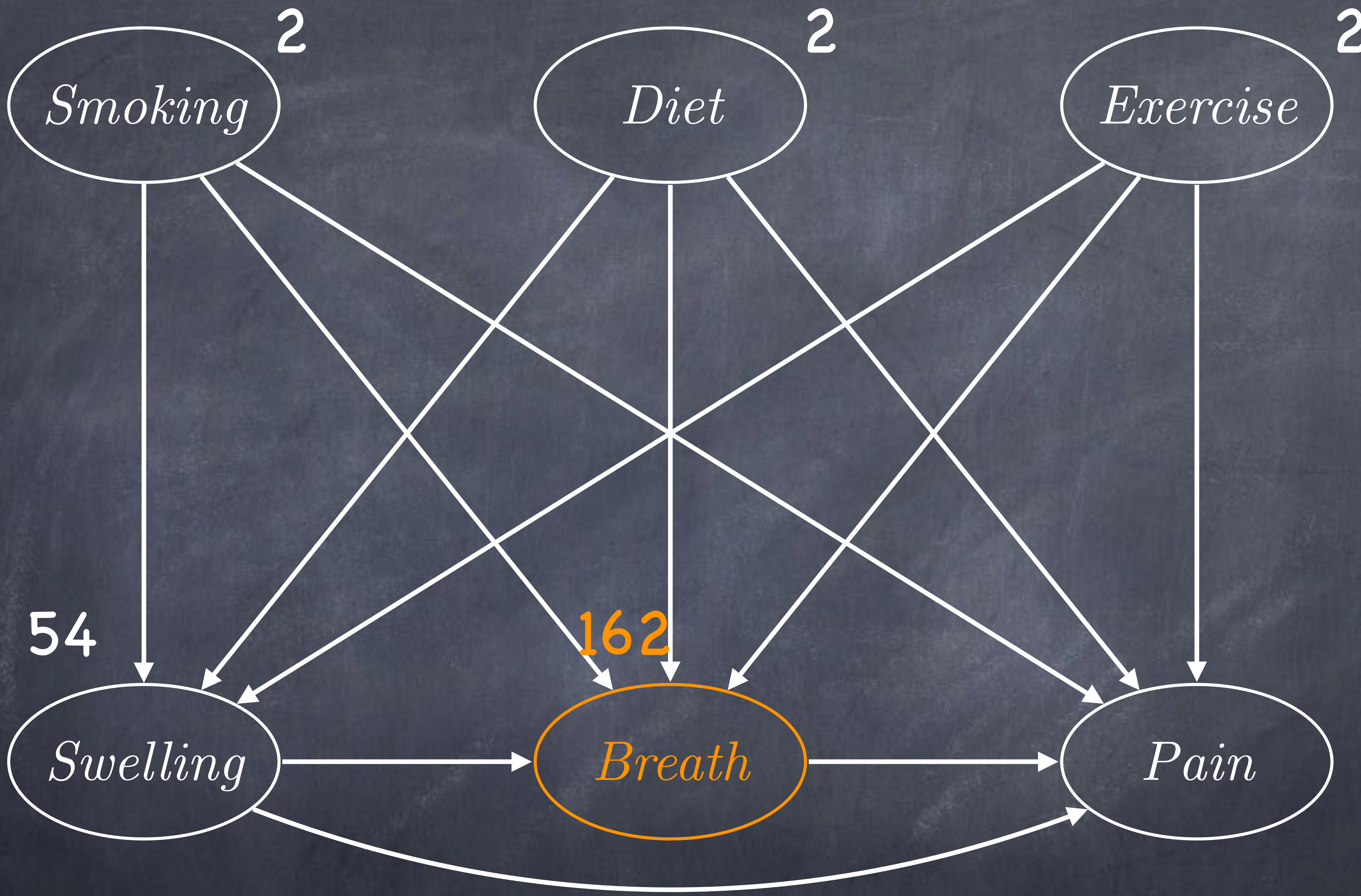
3 values/variable



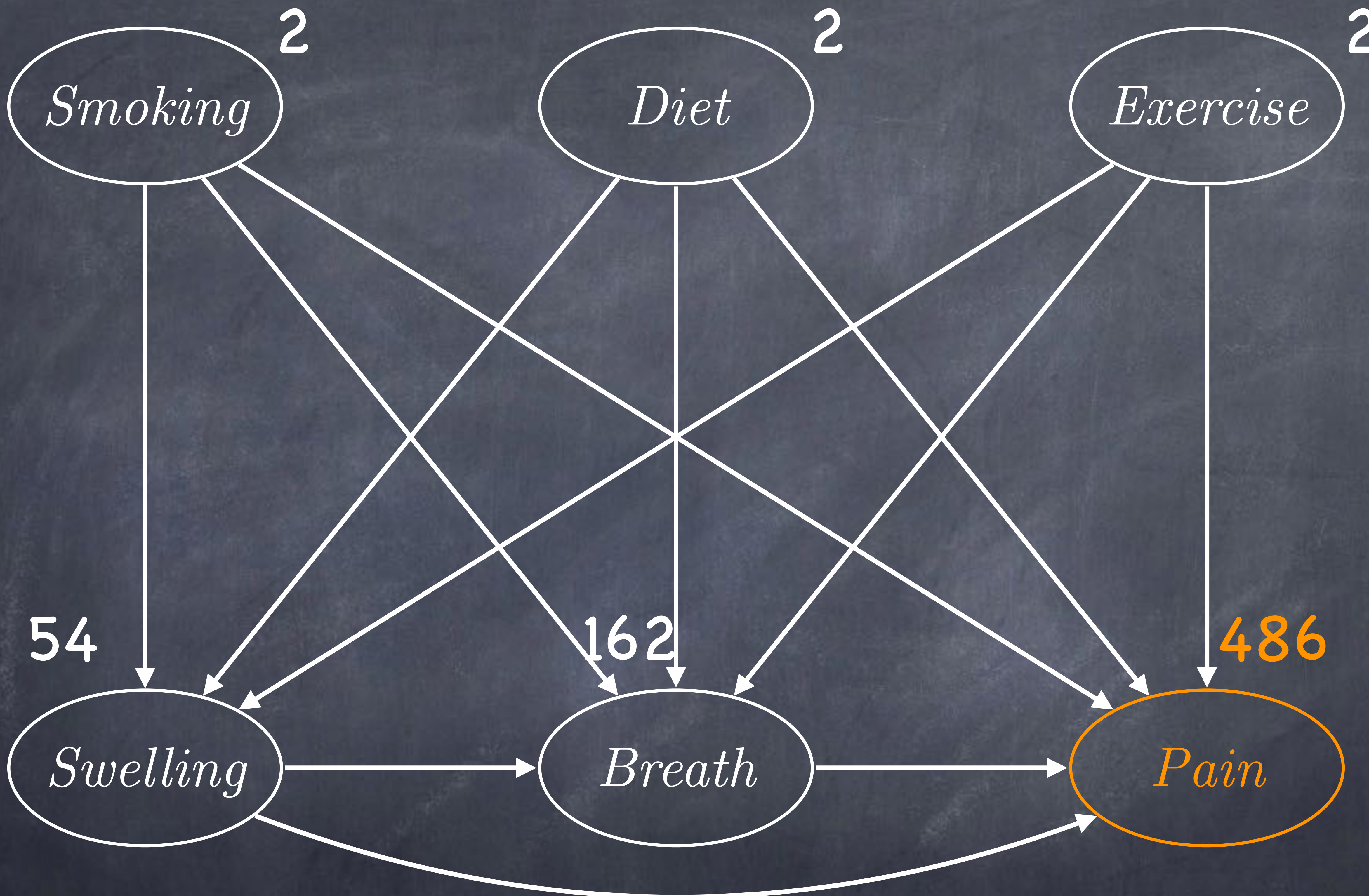
3 values/variable



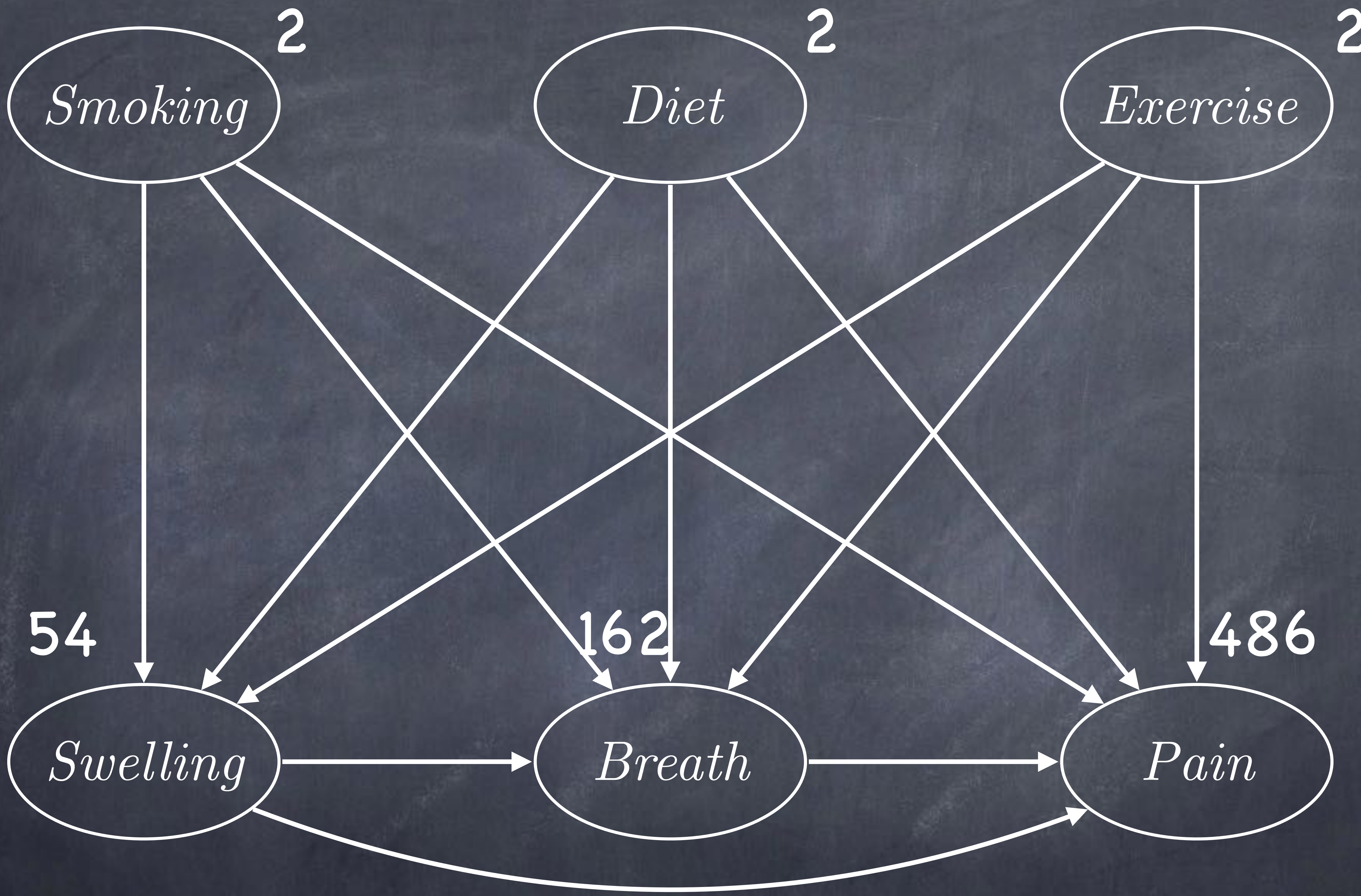
3 values/variable



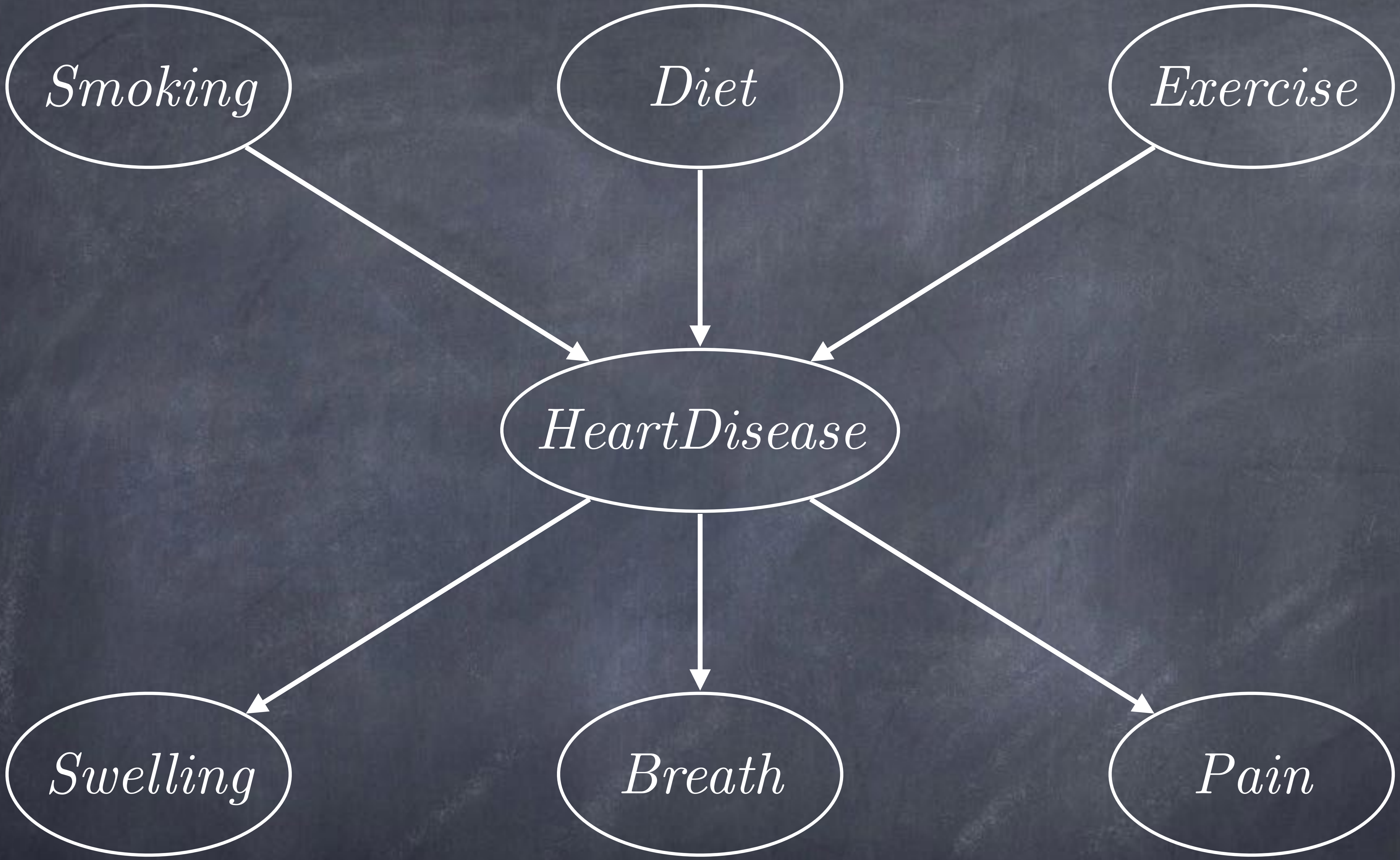
3 values/variable

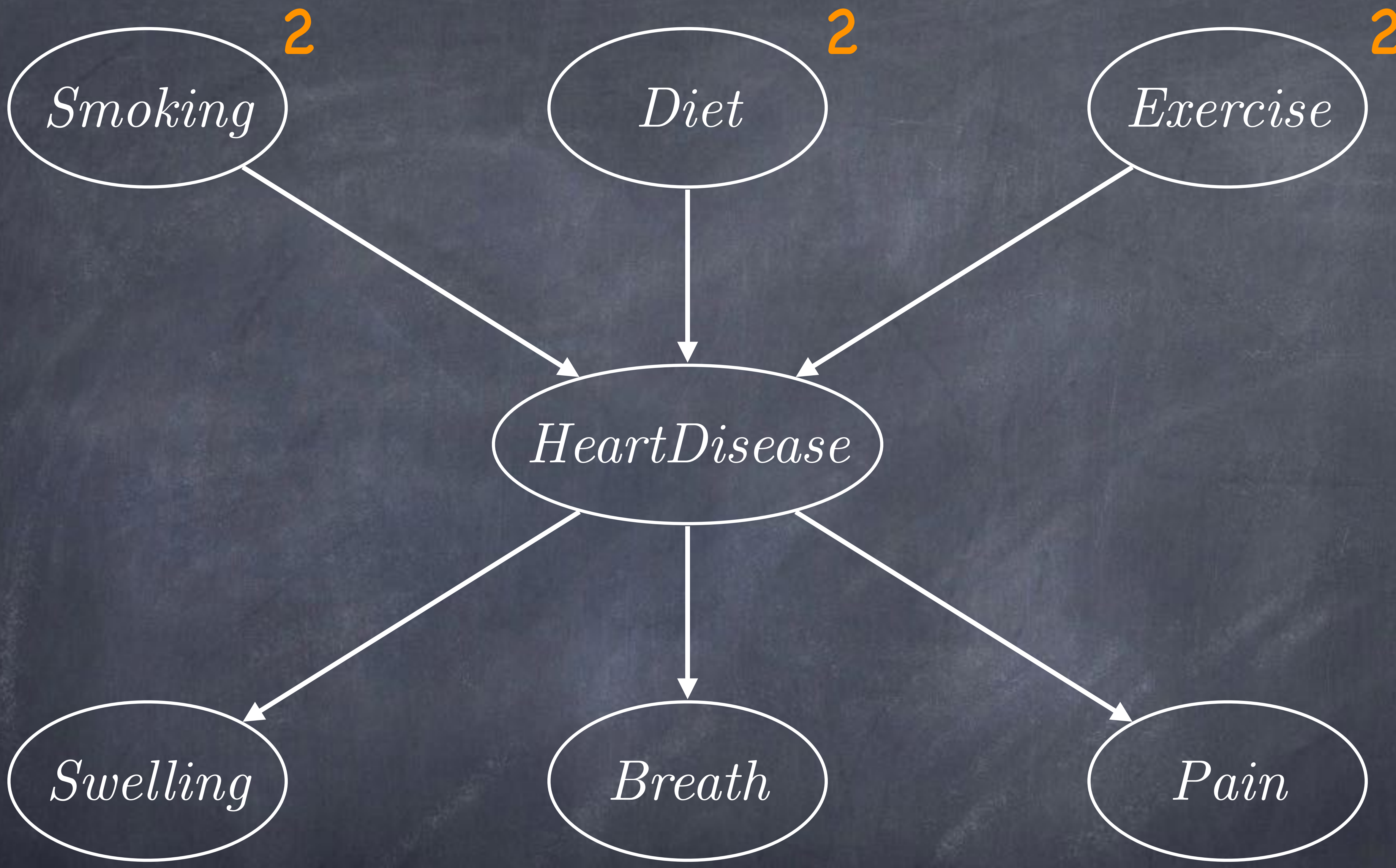


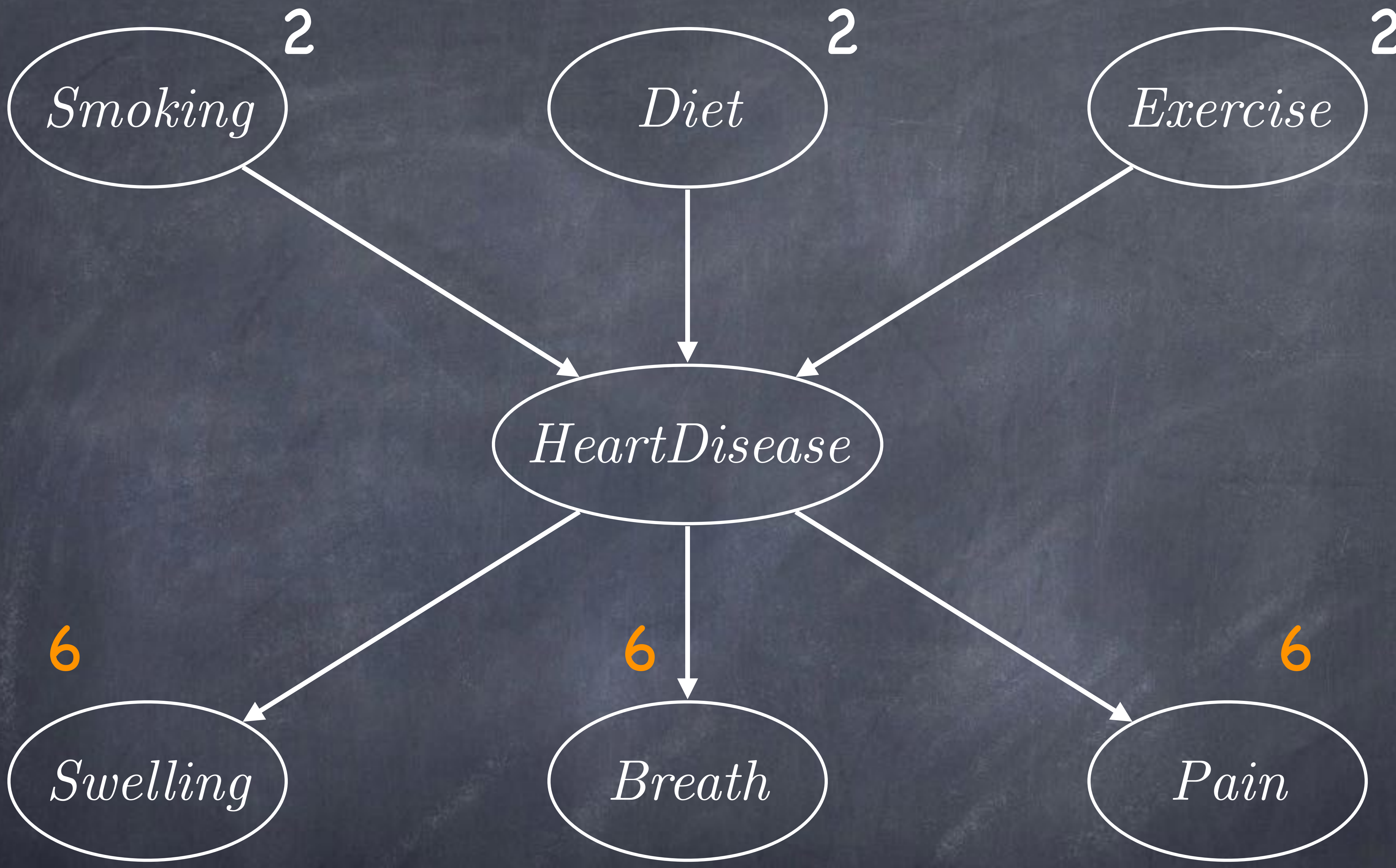
3 values/variable

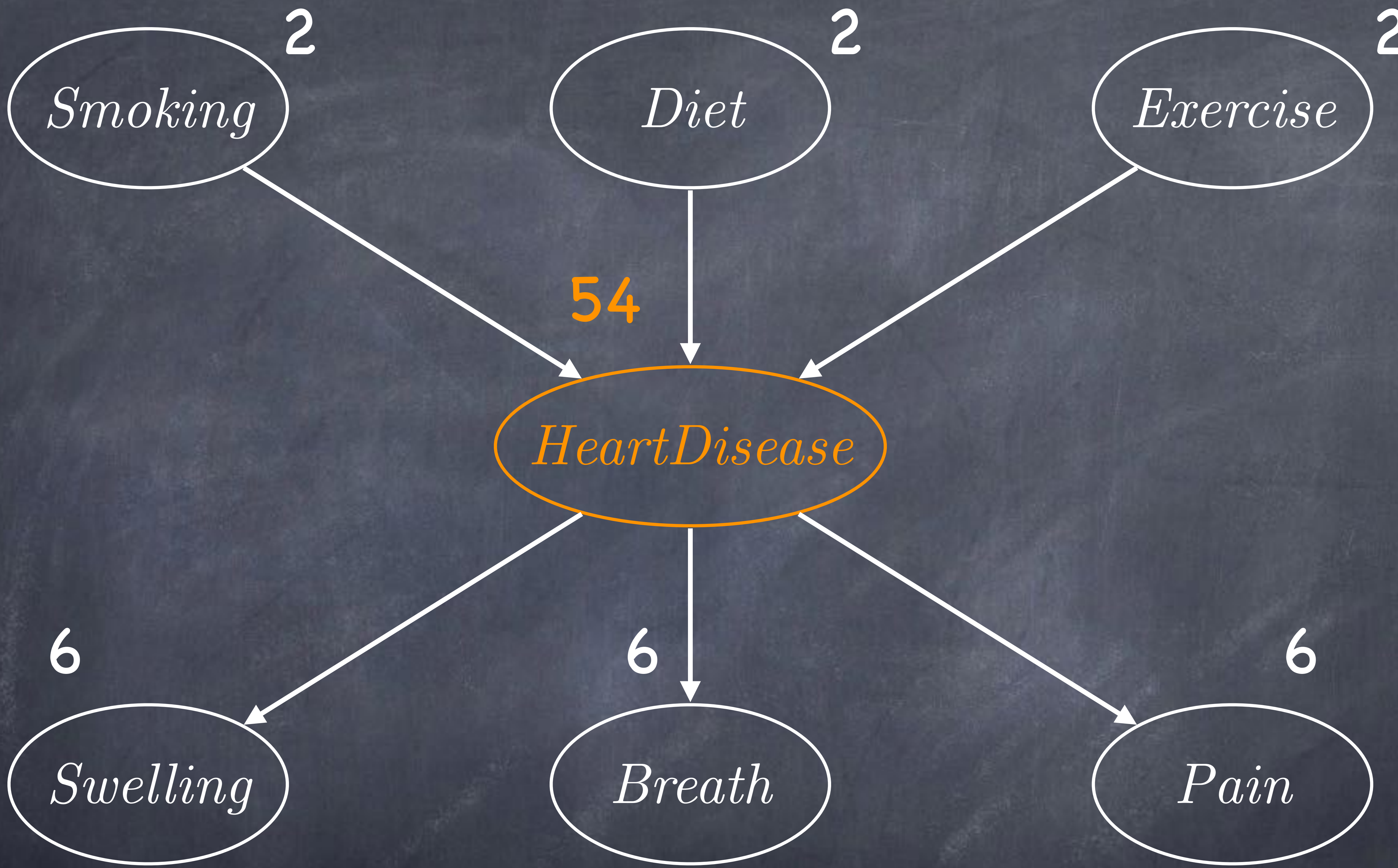


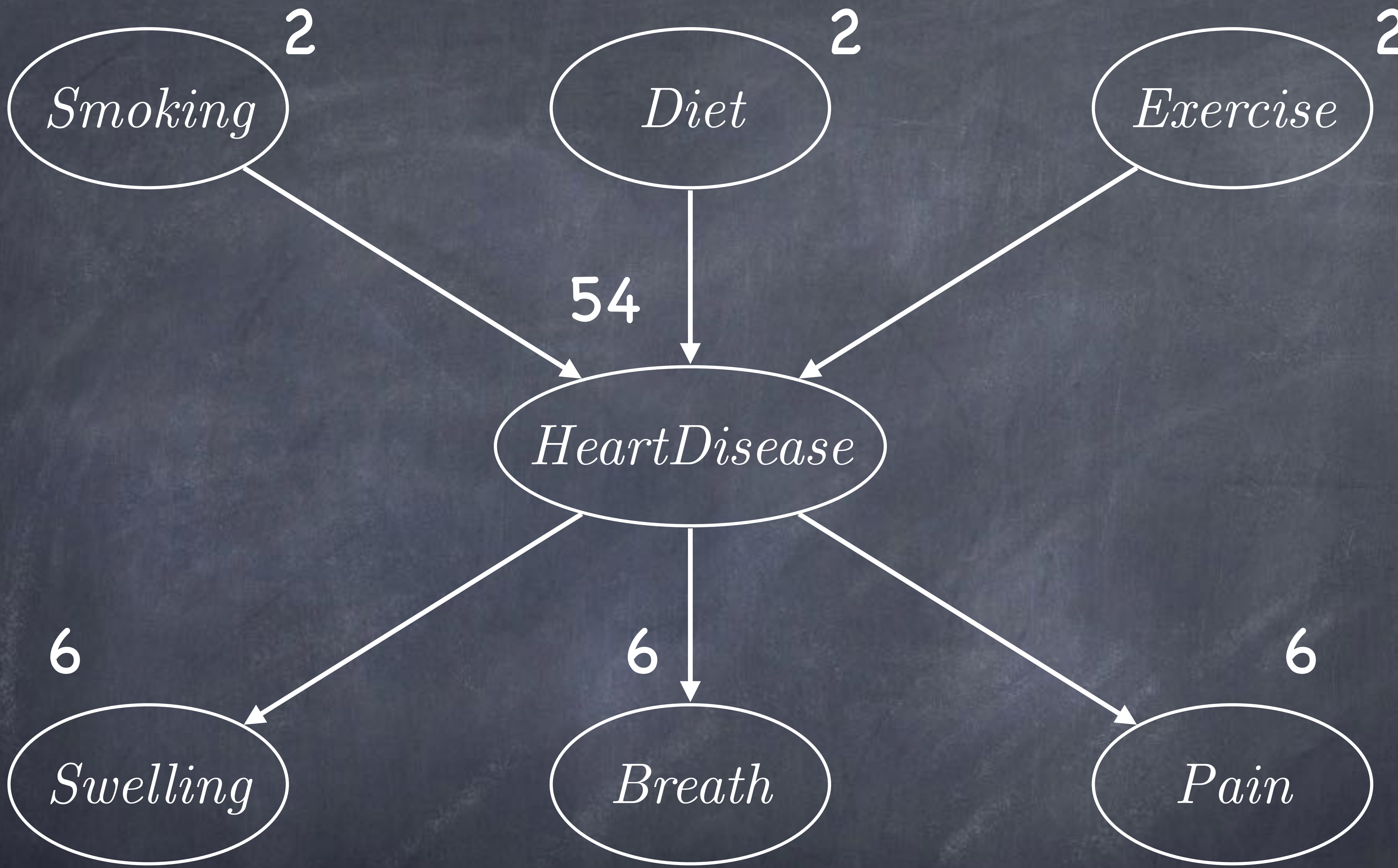
708 parameters!


Smoking
Diet
Exercise
HeartDisease
Swelling
Breath
Pain


2
2
2
Smoking
Diet
Exercise
HeartDisease
Swelling
Breath
Pain


2
2
2
Smoking
Diet
Exercise
HeartDisease
6
6
6
Swelling
Breath
Pain


2
2
2
Smoking
Diet
Exercise
54
HeartDisease
6
6
6
Swelling
Breath
Pain



78 parameters!

$Smoking$	$Diet$	$Exercise$	$Swelling$	$Breath$	$Pain$	$Disease$
low	low	$high$	low	low	low	$?$
$high$	med	$high$	med	$high$	med	$?$
low	low	med	low	low	med	$?$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$?$

Hidden!

Hidden (Latent) Variables

- Can dramatically reduce the number of parameters required to specify a Bayes net
  - Reduces amount of data required to learn the parameters
- Values of hidden variables not present in training data (observations)
  - "Complicates" the learning problem

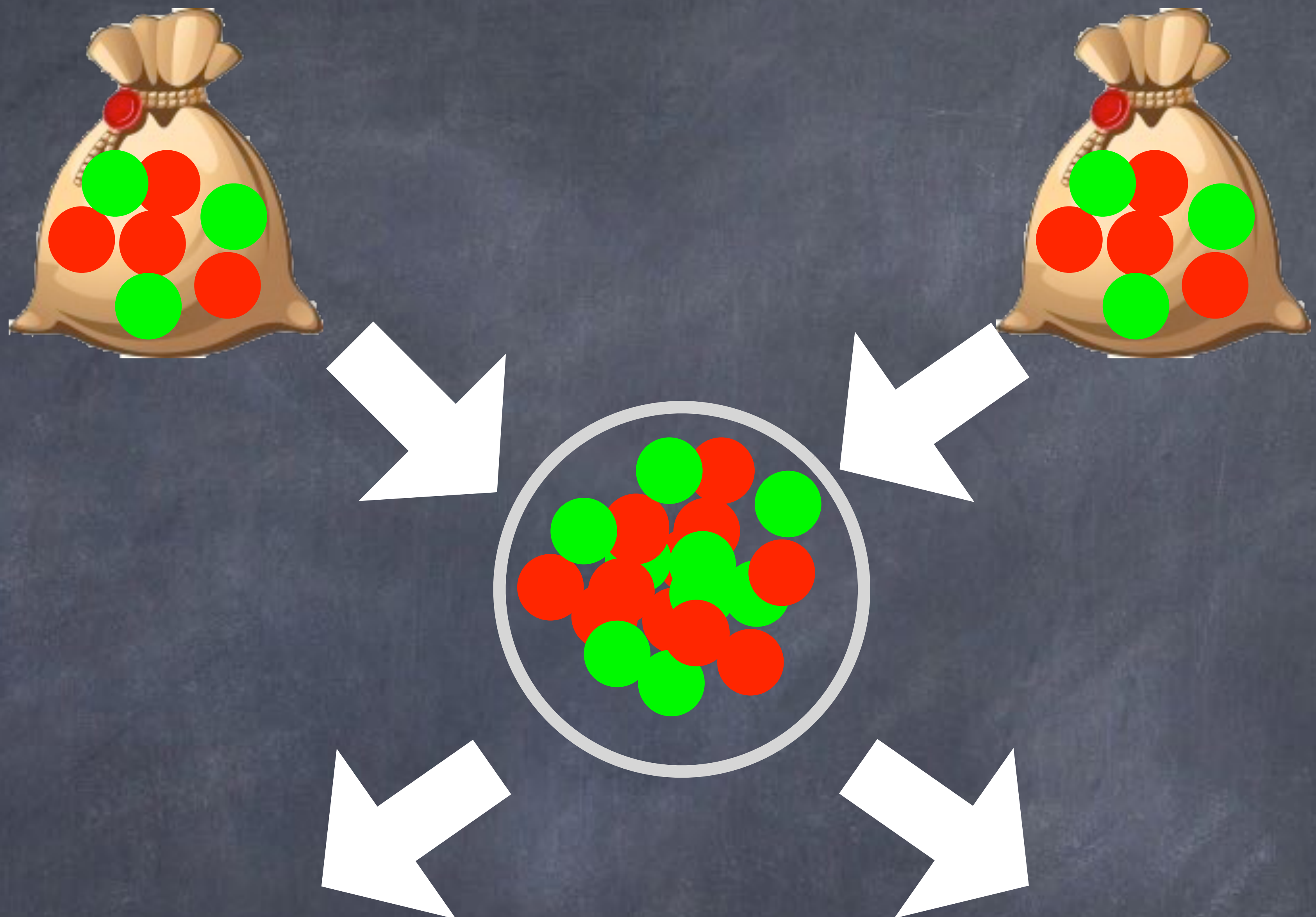
EM
Expectation-Maximization

- Repeat
  - Expectation: “Pretend” we know the parameters and compute (or estimate) likelihood of data given model
  - Maximization: Recompute parameters using expected values as if they were observed values
- Until convergence

Flavor	*cherry*	*lime*
Wrapper	*red*	*green*
Hole	*true*	*false*

$\mathbf{P}(F, W, H)$

Flavor	*Wrapper*	*Hole*	$P(f,w,h)$
cherry	*red*	t	$p_{c,r,t}$
		f	$p_{c,r,f}$
	green	t	$p_{c,g,t}$
		f	$p_{c,g,f}$
lime	*red*	t	$p_{l,r,t}$
		f	$p_{l,r,f}$
	green	t	$p_{l,g,t}$
		f	$p_{l,g,f}$



$\mathbf{P}_1(F, W, H)$ $\mathbf{P}_2(F, W, H)$

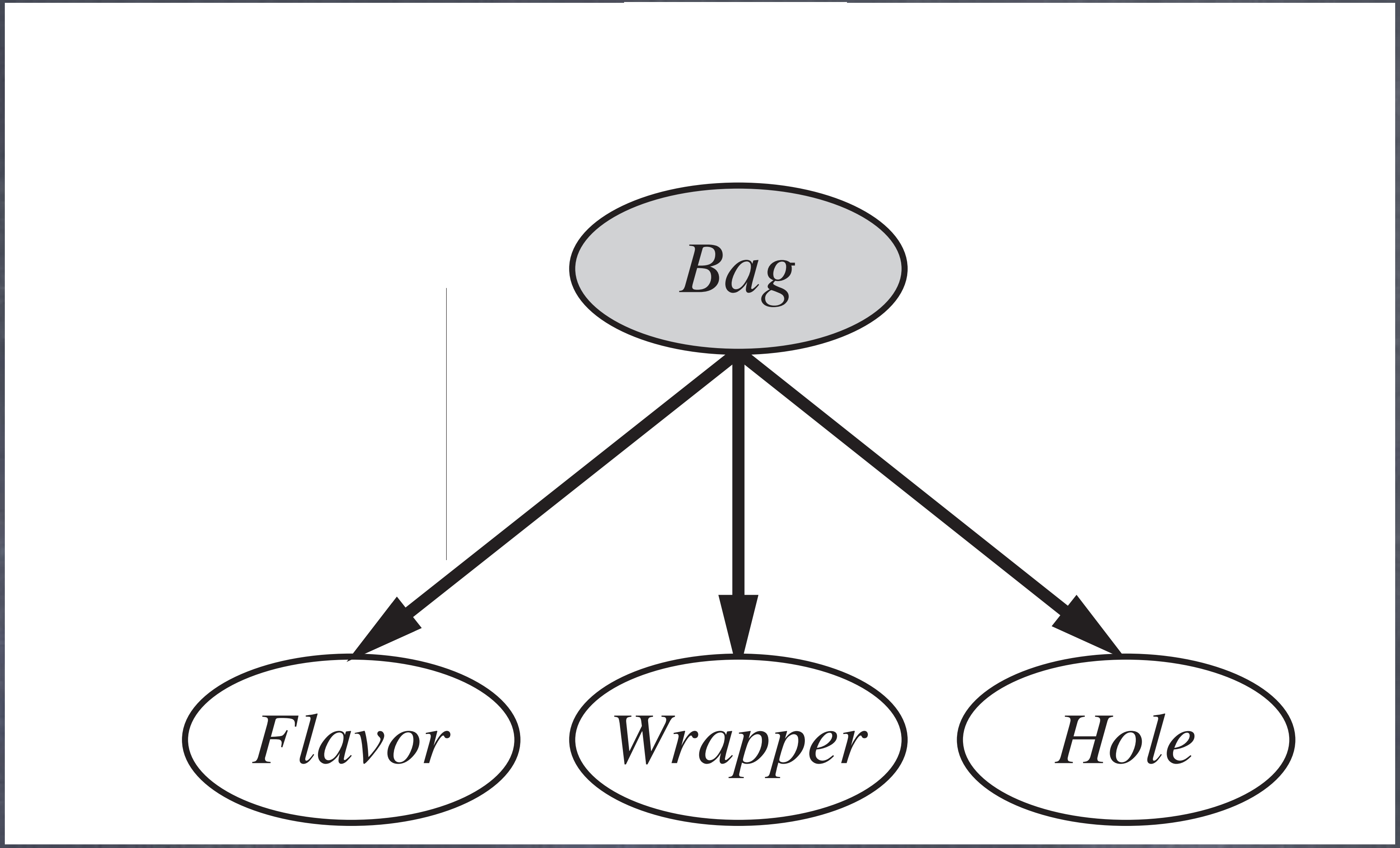
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam euismod euismod facilisis. Aliquam erat volutpat. Maecenas nisl ligula, dignissim et volutpat ac, pharetra blandit augue. Maecenas id ligula in leo tristique viverra. Curabitur lacinia nulla in nibh bibendum laoreet. Morbi a est mi, mattis imperdiet risus. Quisque quam felis, facilisis ac semper vel, viverra vitae nulla. Donec nisl lectus, faucibus vehicula tincidunt nec, ultrices nec eros. Proin non felis nec urna pellentesque tempor at sit amet est.

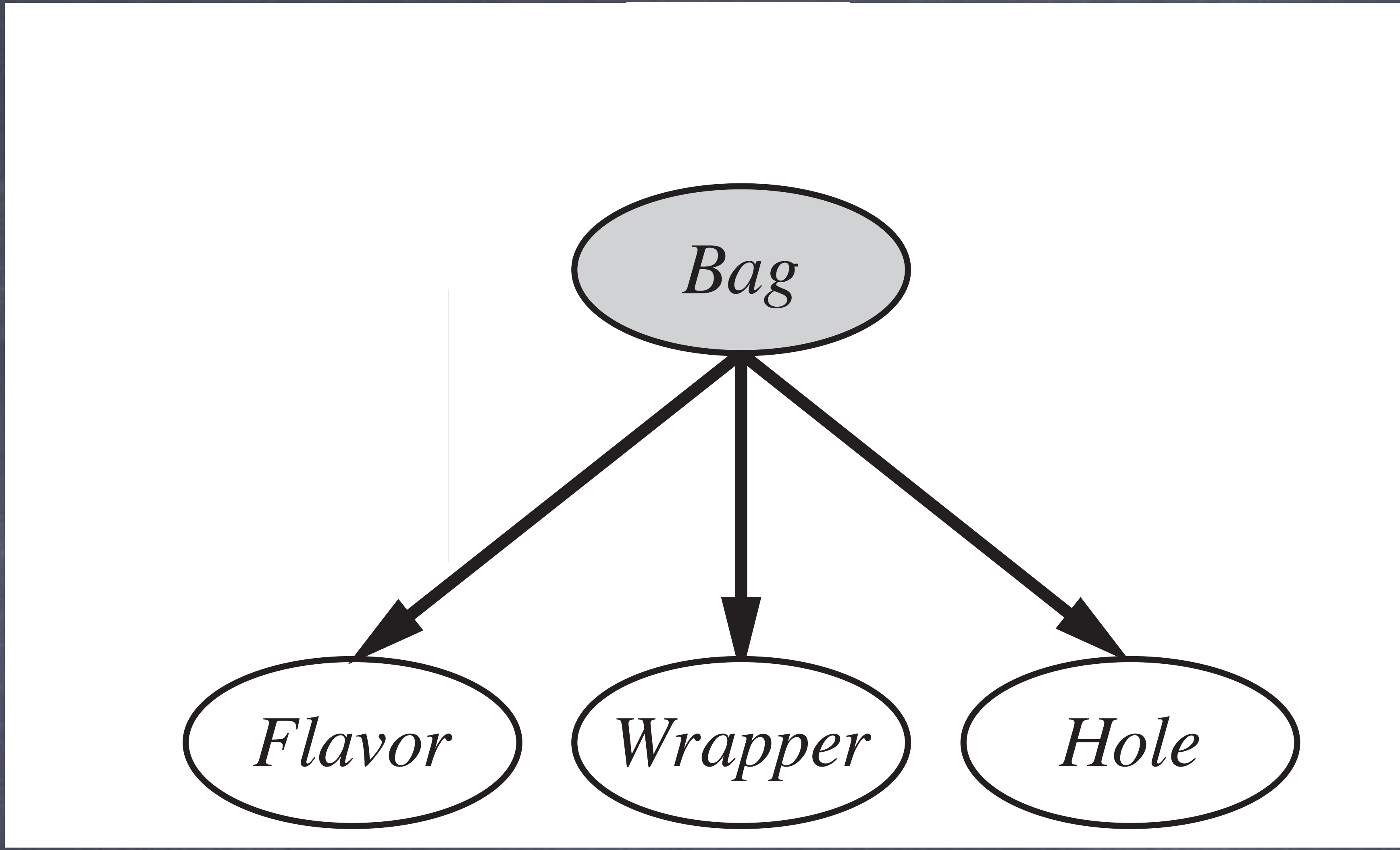
$\mathbf{P}_1(X_1, X_2, X_3)$

$\mathbf{P}_2(X_1, X_2, X_3)$

$\mathbf{P}(F, W, H)$

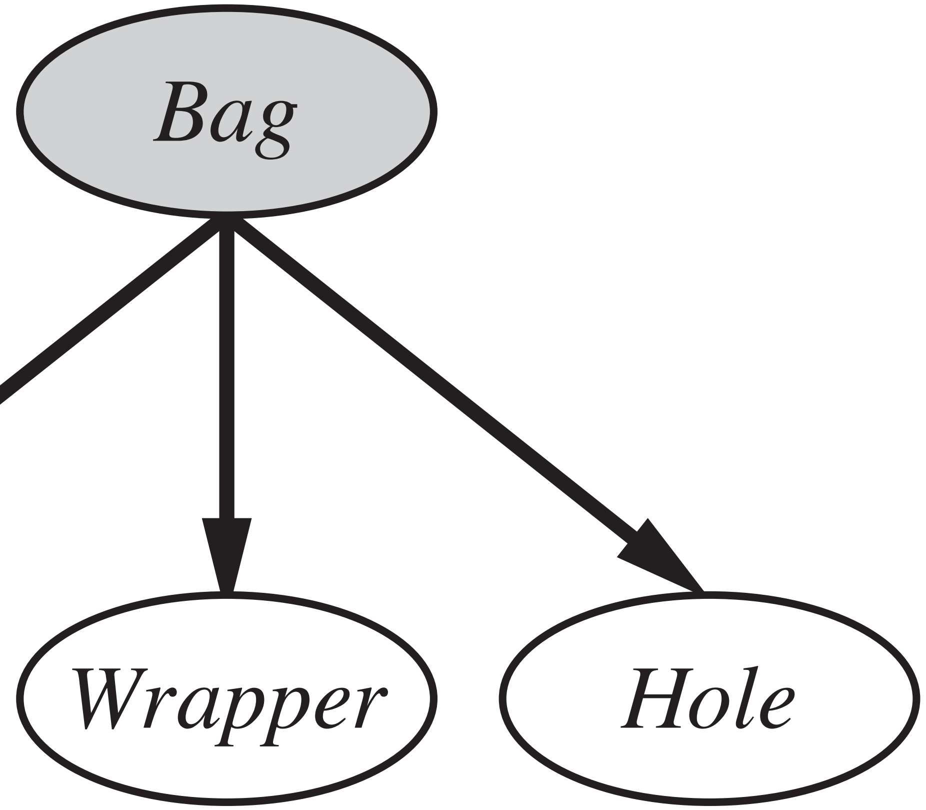
$Flavor$	$Wrapper$	$Hole$	$P(f, w, h)$
$cherry$	red	t	$p_{c,r,t}$
		f	$p_{c,r,f}$
	$green$	t	$p_{c,g,t}$
		f	$p_{c,g,f}$
$lime$	red	t	$p_{l,r,t}$
		f	$p_{l,r,f}$
	$green$	t	$p_{l,g,t}$
		f	$p_{l,g,f}$


Bag
Flavor
Wrapper
Hole



Bag	$P(W{=}red \mid B)$
1	Θ_{W1}
2	Θ_{W2}

Bag	$P(H{=}true \mid B)$
1	Θ_{H1}
2	Θ_{H2}



$P(Bag{=}1)$
θ

Bag	$P(F{=}cherry \mid B)$
1	θ_{F1}
2	θ_{F2}

Bag	$P(W{=}red \mid B)$
1	Θ_{W1}
2	Θ_{W2}

Bag	$P(H{=}true \mid B)$
1	Θ_{H1}
2	Θ_{H2}

7 parameters

Flavor	Wrapper	Hole	Bag
cherry	red	true	?
cherry	red	true	?
lime	green	false	?
cherry	green	true	?
lime	green	true	?
cherry	red	false	?
lime	red	true	?

Hidden!

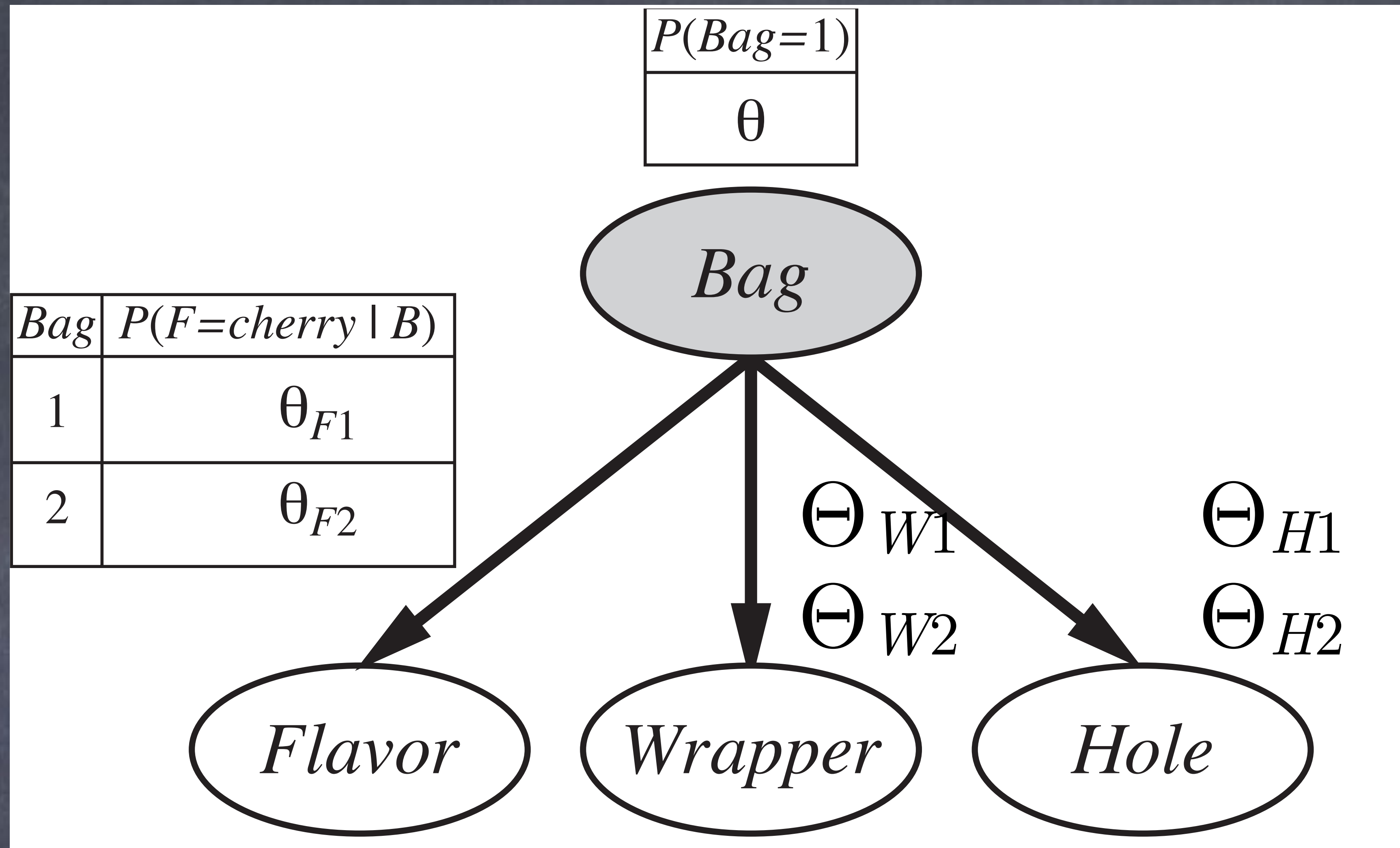
$N=1000$	$W=red$		$W=green$	
	$H=yes$	$H=no$	$H=yes$	$H=no$
$F=cherry$	273	93	104	90
$F=lime$	79	100	94	167

$N=1000$	$W=red$		$W=green$	
	$H=yes$	$H=no$	$H=yes$	$H=no$
$F=cherry$	273	93	104	90
$F=lime$	79	100	94	167

No values for Bag

EM
Expectation-Maximization

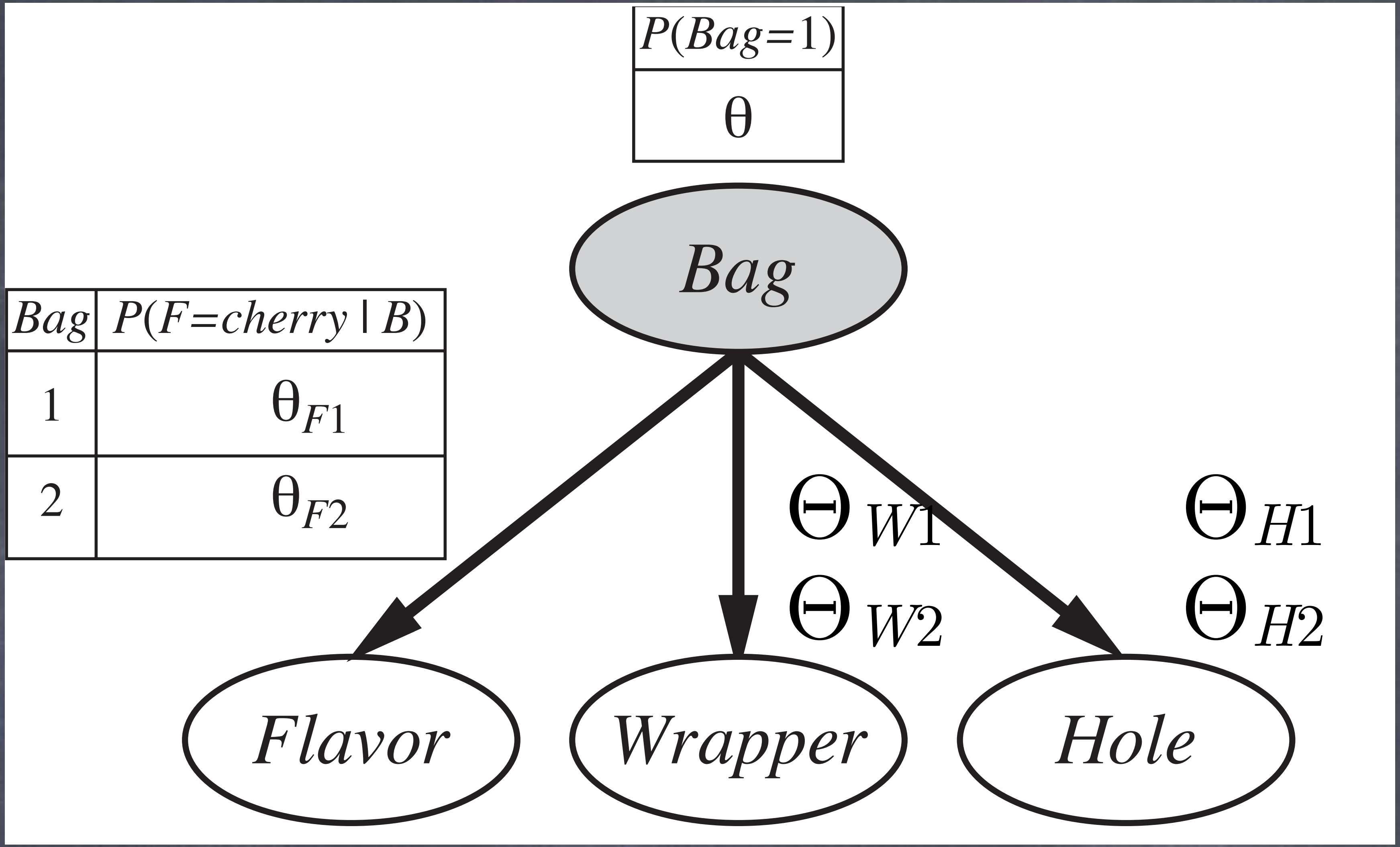
- Repeat
    - E: Use the current values of the parameters to compute the expected values of the hidden variables
    - M: Recompute the parameters to maximize the likelihood of the data given the values of (all) the variables
- Until convergence



$$\Theta = N(B=1)/N$$

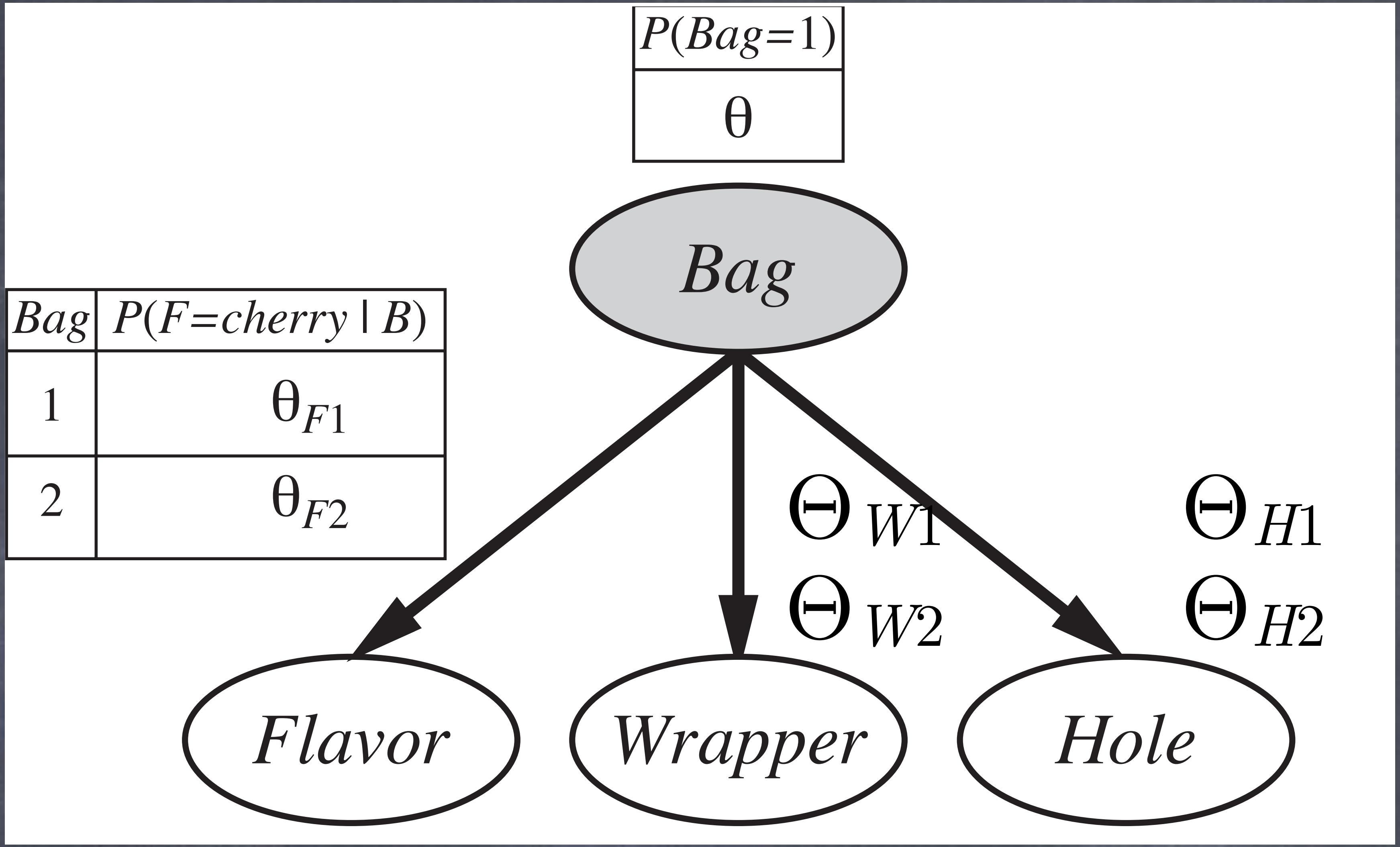
$$\Theta_{F1}, \Theta_{W1}, \Theta_{H1}$$

$$\Theta_{F2}, \Theta_{W2}, \Theta_{H2}$$

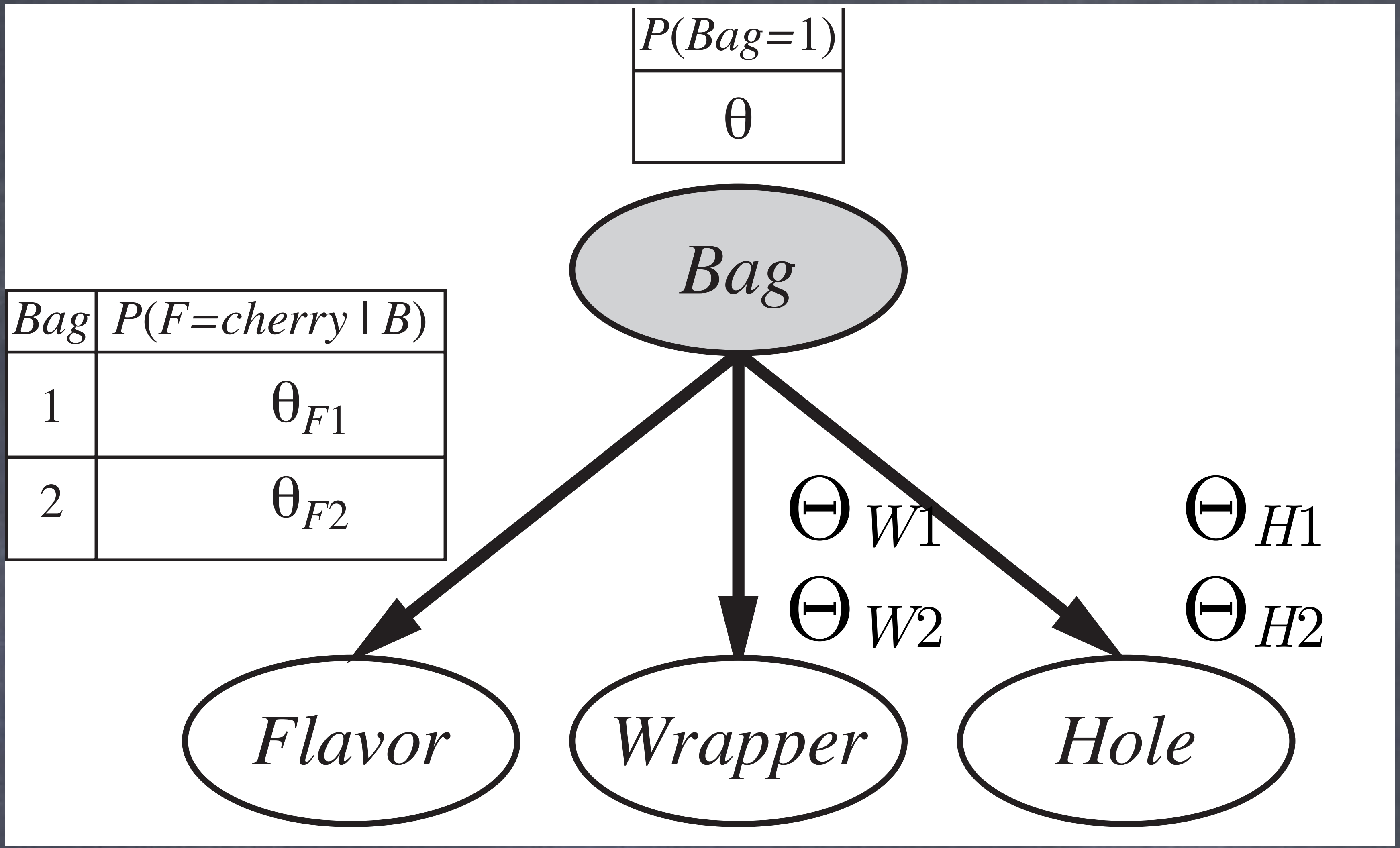


We don't know $N(B{=}1)$

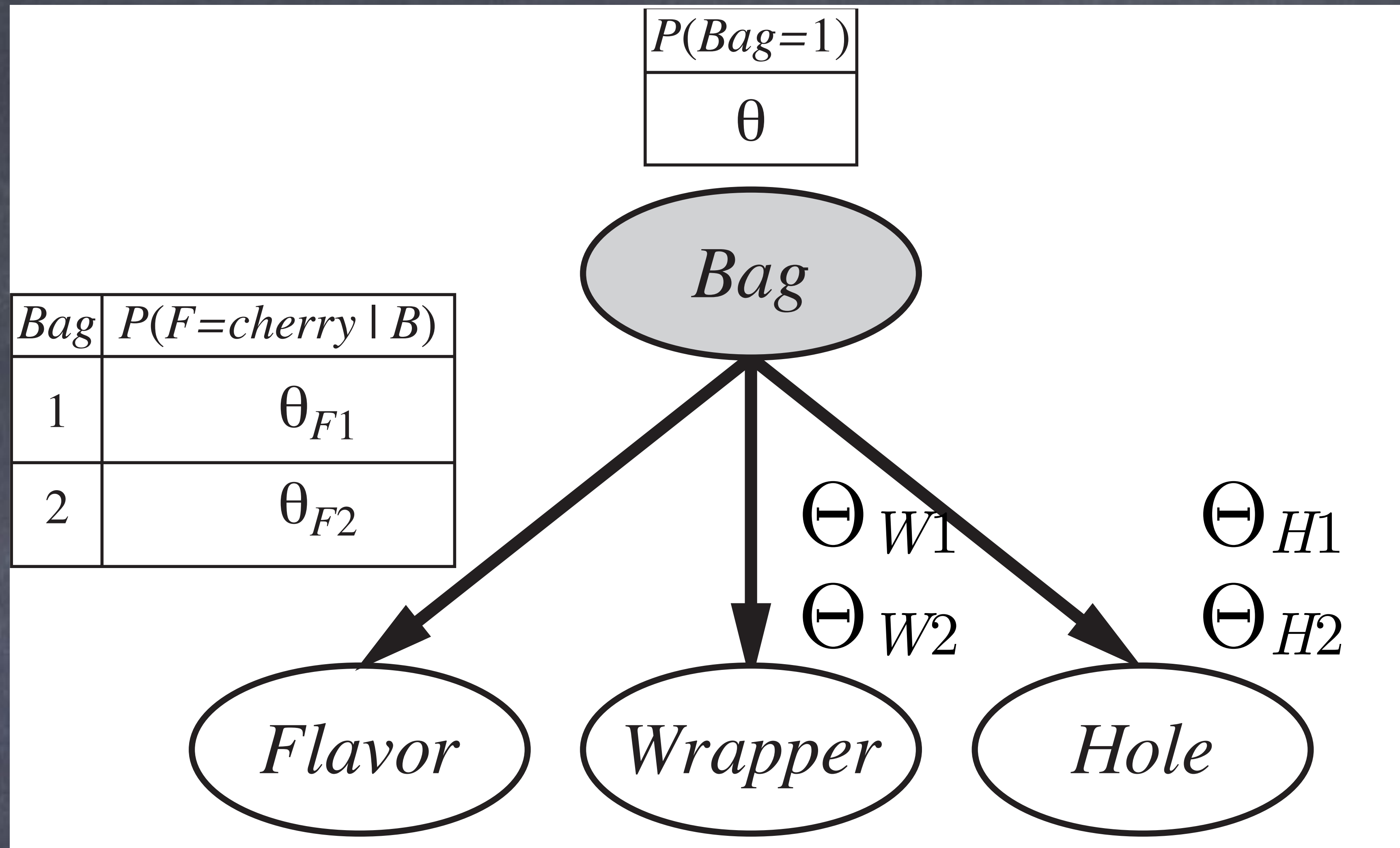
Estimate: $\hat{N}(B{=}1)$



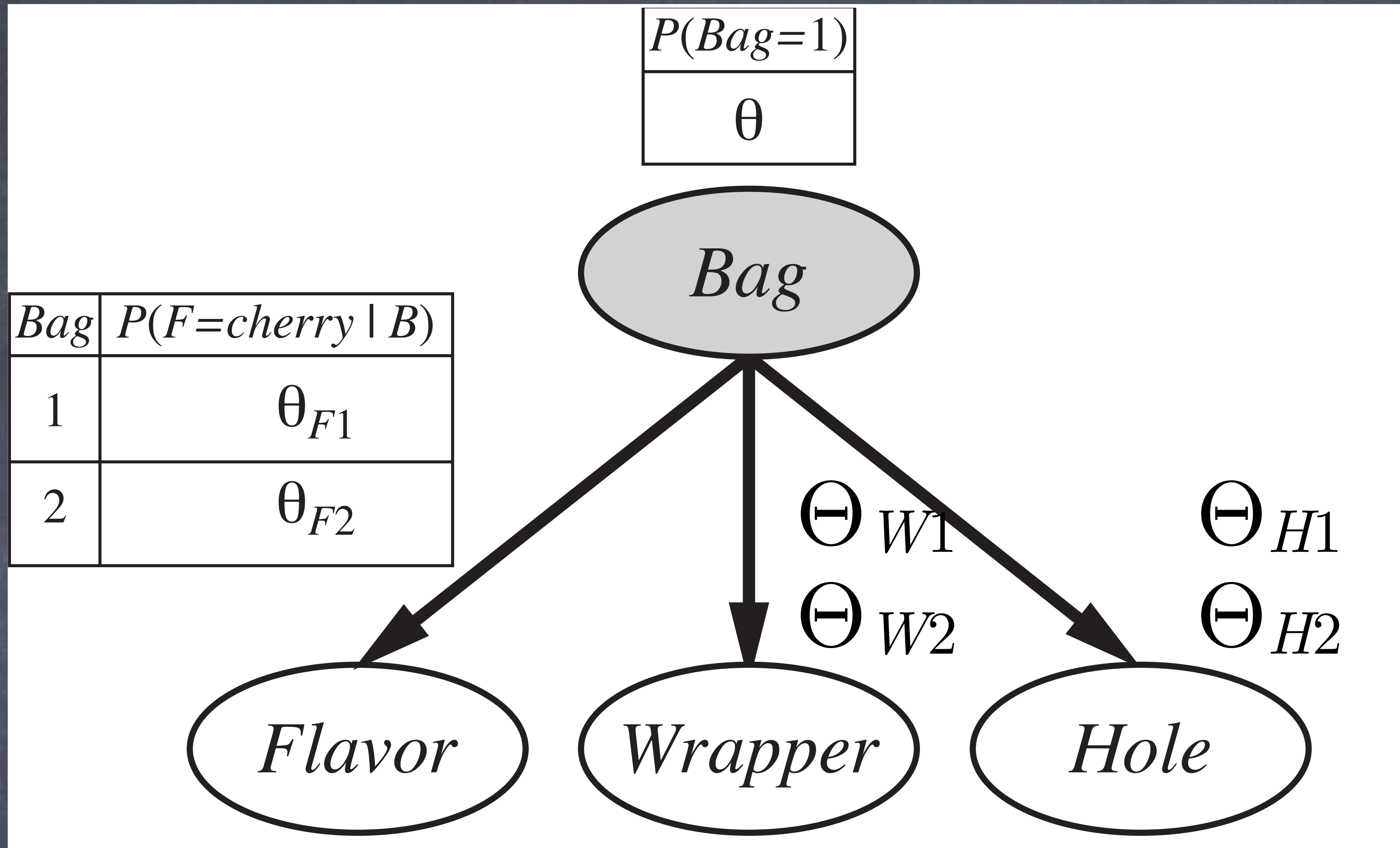
$$Flavor\!=\!cherry,\ Wrapper\!=\!red,\ Hole\!=\!true$$



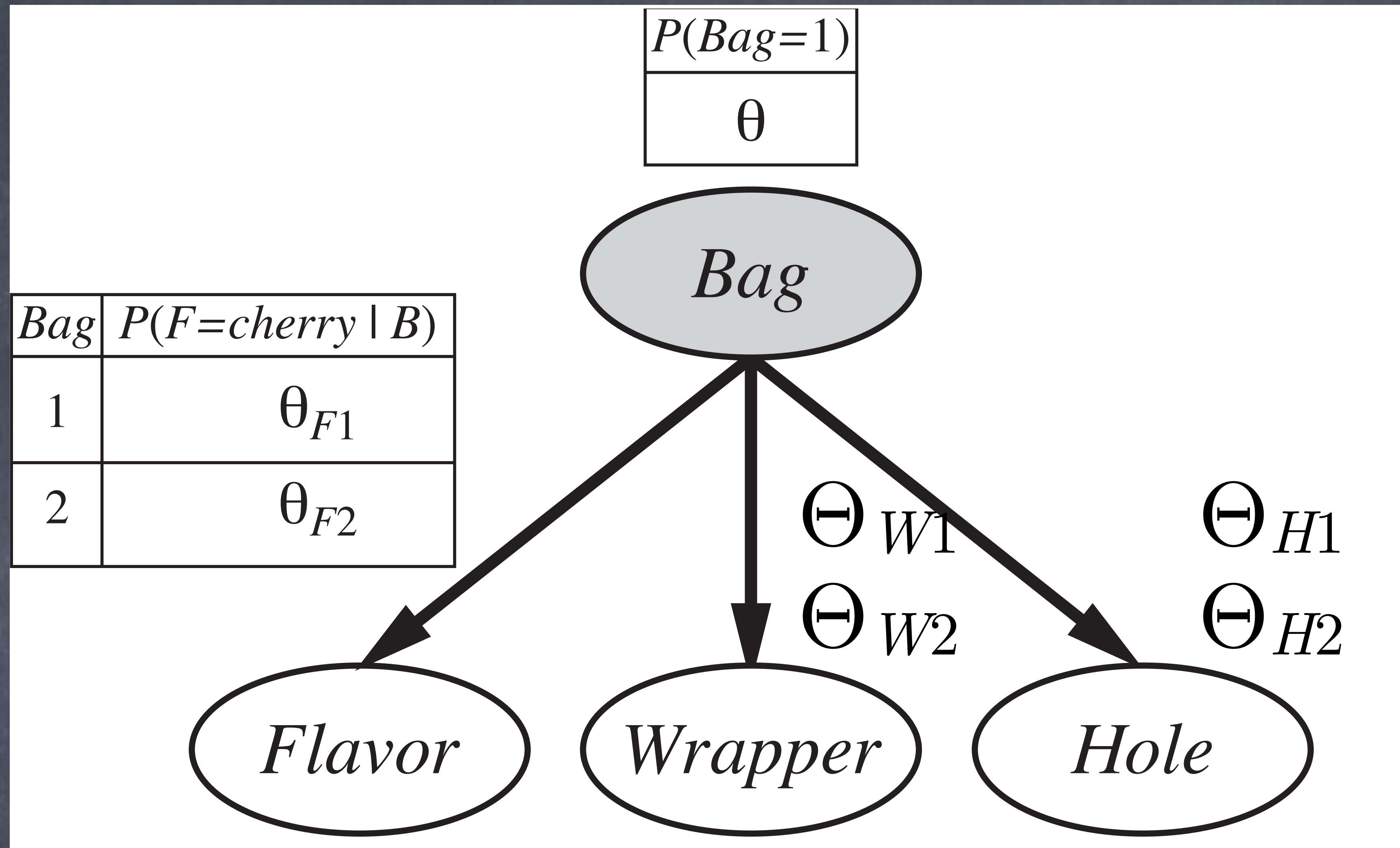
$$P(Bag{=}1 \mid Flavor{=}cherry,\ Wrapper{=}red,\ Hole{=}true)?$$



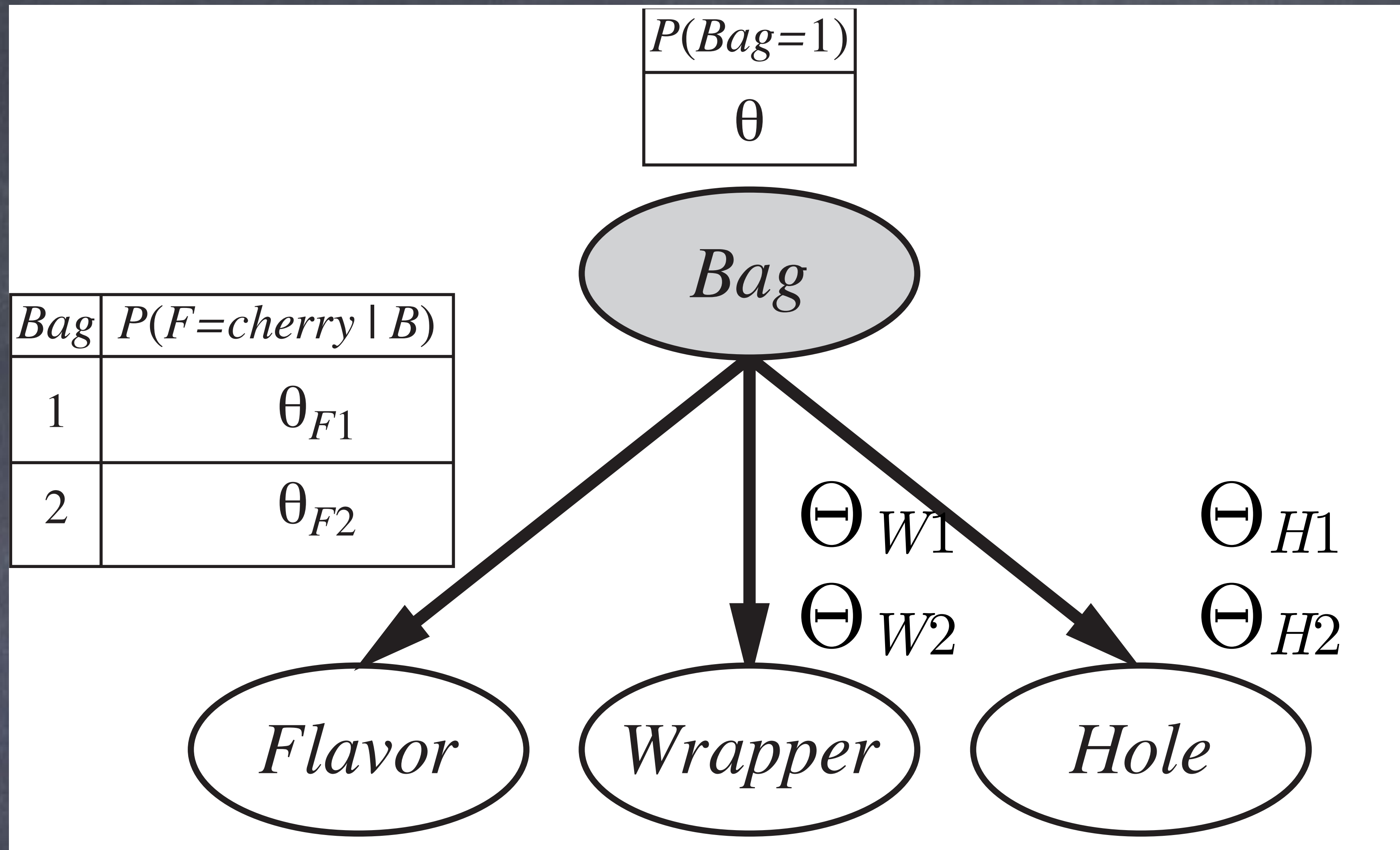
$$
\begin{aligned}
&P(B=1 \mid F=c, W=r, H=t) \\
&= \alpha P(B=1, F=c, W=r, H=t) \\
&= \alpha P(B=1) P(F=c \mid B=1) \\
&\qquad\qquad P(W=r \mid B=1) P(H=t \mid B=1) \\
&= \alpha \Theta \, \Theta_{F1} \, \Theta_{W1} \, \Theta_{H1}
\end{aligned}
$$



$$\hat{N}(B=1) = \sum_{j=1}^{N} \alpha P(B=1 \mid F=f_j, W=w_j, H=h_j)$$



$$\Theta \leftarrow \hat{N}(B{=}1)/\mathrm{N}$$



$$\Theta \leftarrow \hat{N}(B{=}1)/\mathrm{N}$$

$$\Theta_{F1},\ \Theta_{W1},\ \Theta_{H1}$$

$$\Theta_{F2},\ \Theta_{W2},\ \Theta_{H2}$$

$\Theta^{(0)} = 0.6$

$\Theta^{(0)}_{F1} = \Theta^{(0)}_{W1} = \Theta^{(0)}_{H1} = 0.6$

$\Theta^{(0)}_{F2} = \Theta^{(0)}_{W2} = \Theta^{(0)}_{H2} = 0.4$

$\Theta^{(1)} = 0.6124$

$\Theta^{(1)}_{F1} = 0.6684$

$\Theta^{(1)}_{W1} = 0.6483$

$\Theta^{(1)}_{H1} = 0.6558$

$\Theta^{(1)}_{F2} = 0.3887$

$\Theta^{(1)}_{W2} = 0.3817$

$\Theta^{(1)}_{H2} = 0.3827$

$L(\mathbf{d}|h) \approx$ -2044

Actual:

$\Theta = 0.5$

$\Theta_{F1} = \Theta_{W1} = \Theta_{H1} = 0.8$

$\Theta_{F2} = \Theta_{W2} = \Theta_{H2} = 0.3$

$L(\mathbf{d}|h) \approx$ -2021

EM
Expectation-Maximization

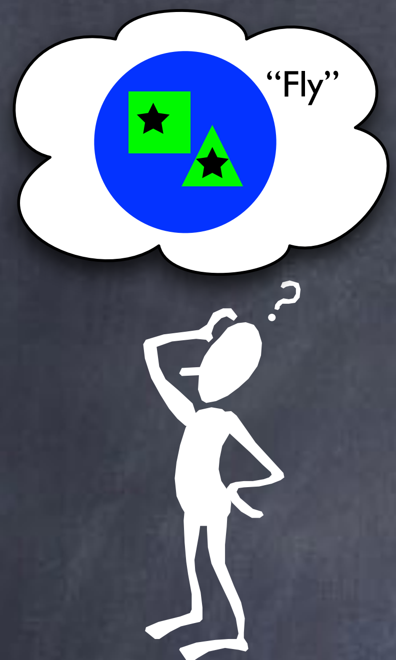
- Repeat
  - E: Pretend we know the values of the parameters and infer the expected values of the hidden variables
  - M: Update the parameters to maximize the likelihood of the data given the values of (all) the variables
- Until convergence

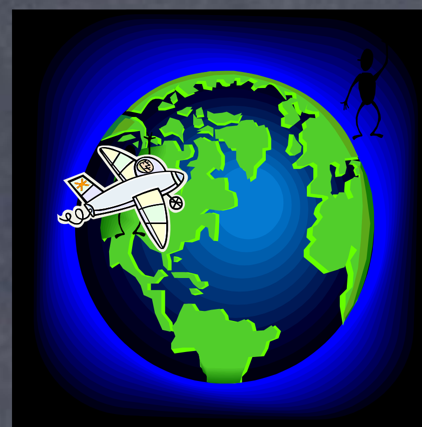
Learning Probabilistic Models

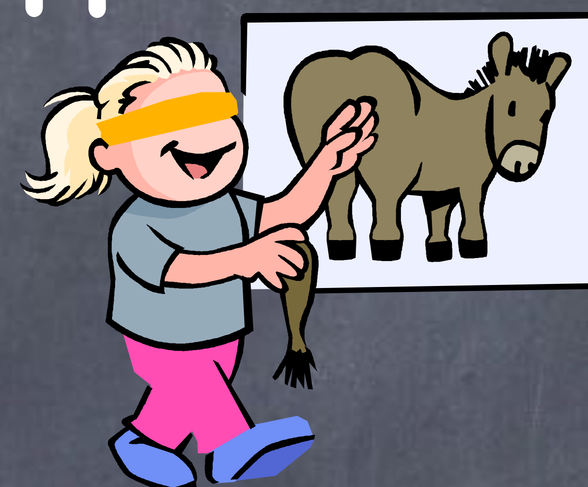
- Learning parameters of probability distributions
- Maximum Likelihood Hypothesis: maximizes the likelihood of the data
- Learning parameters of Bayes Nets
- Naive Bayes classifiers
- EM: Learning with incomplete data

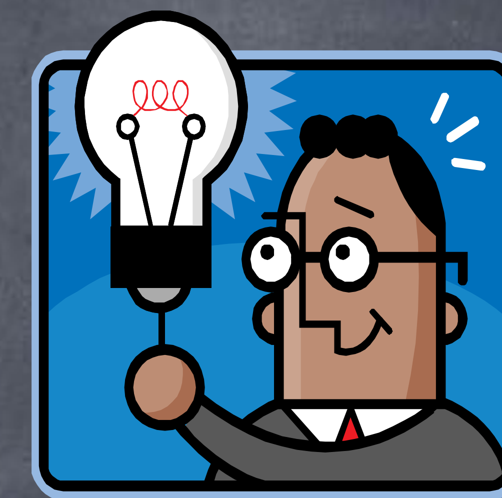
Intro to AI

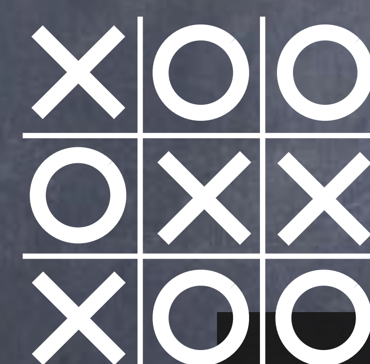
Search

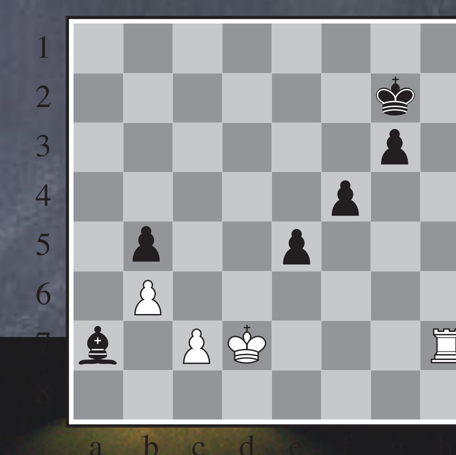


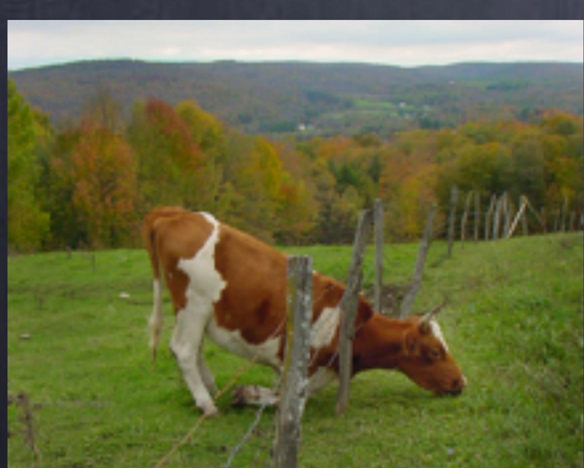


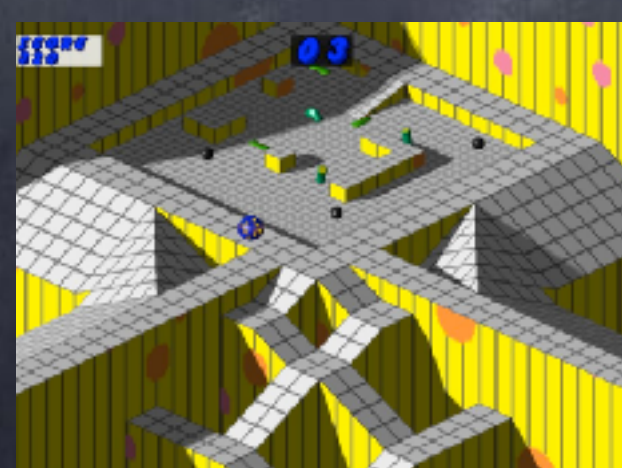




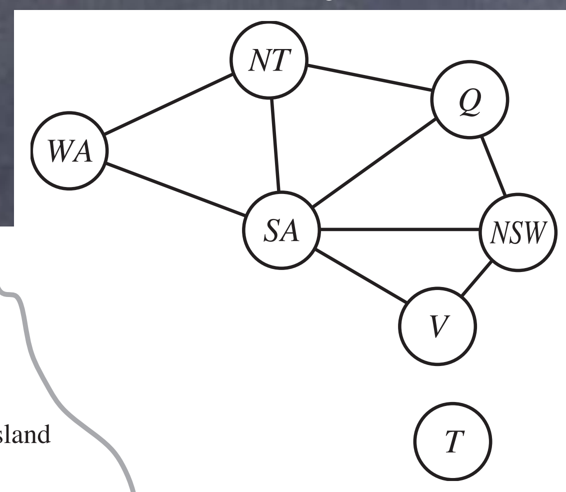


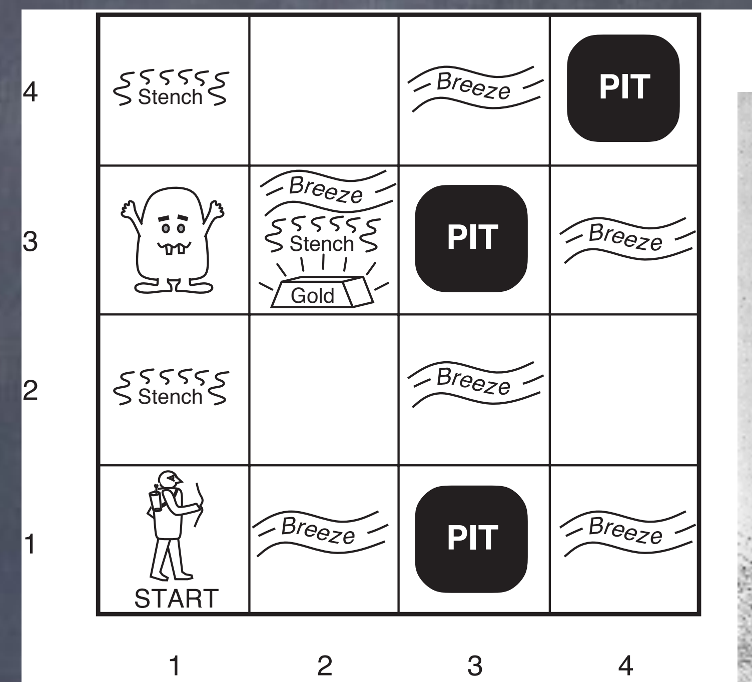


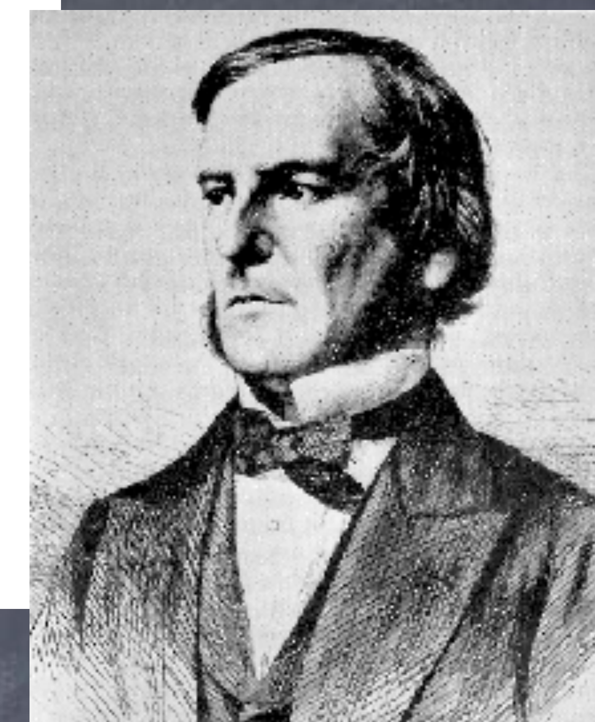


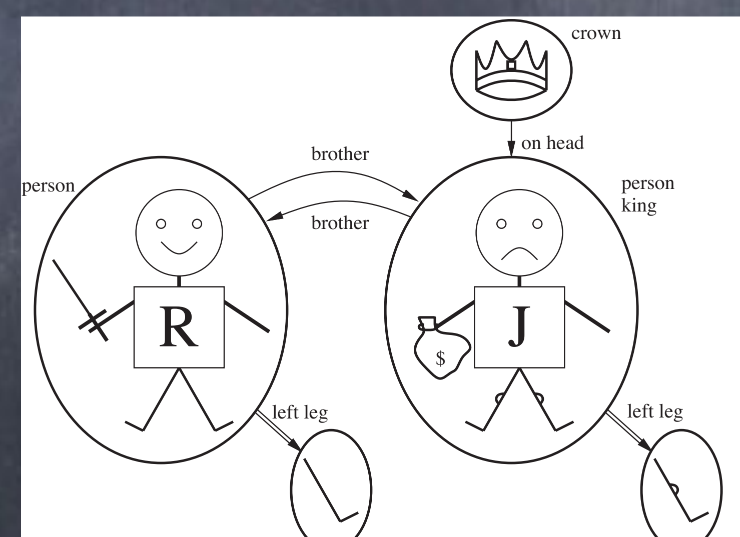


Representation









$$Hungry \lor Cranky$$

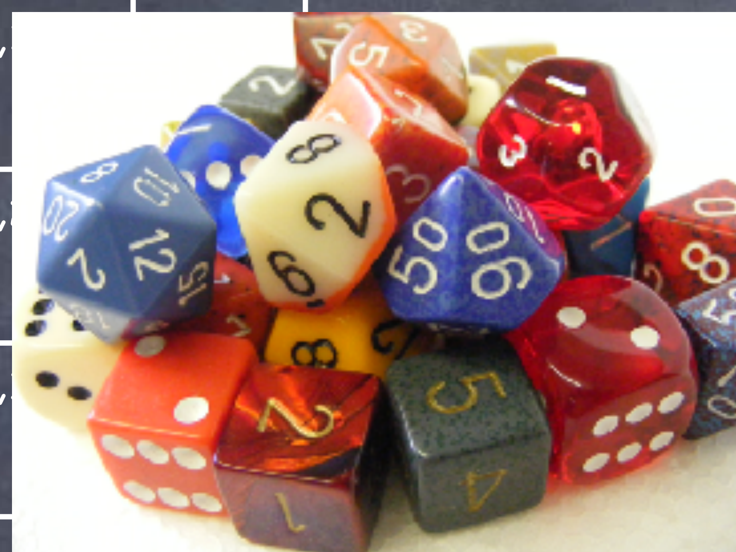
$$\neg Hungry$$

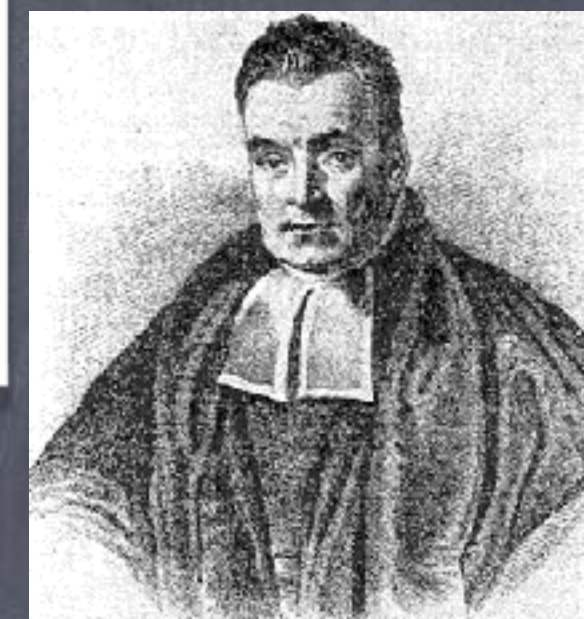
$$Cranky$$

if $\alpha \vdash \beta$ then $\alpha \models \beta$

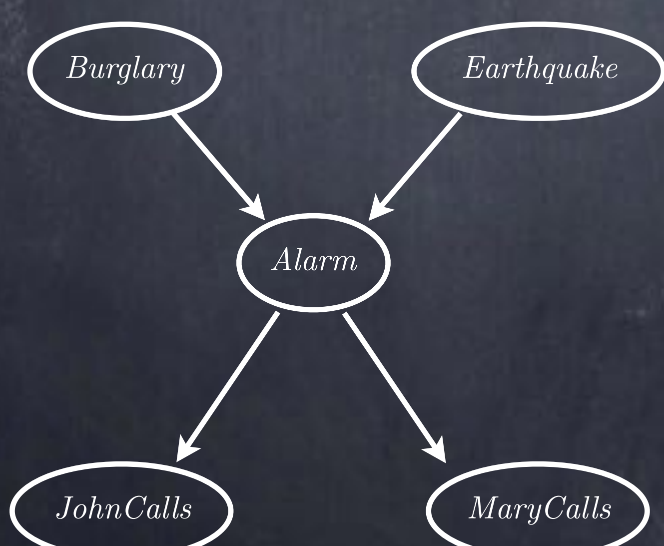
if $\alpha \models \beta$ then $\alpha \vdash \beta$

Uncertainty

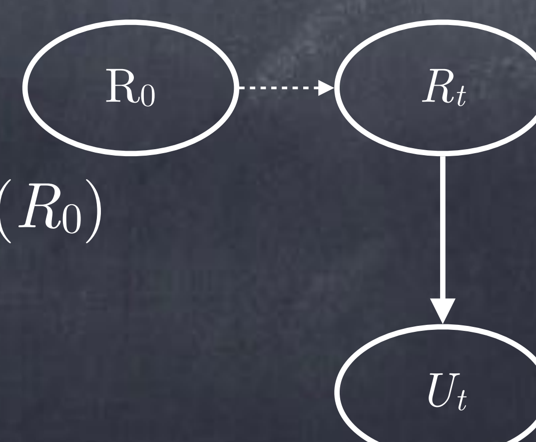




$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$



$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

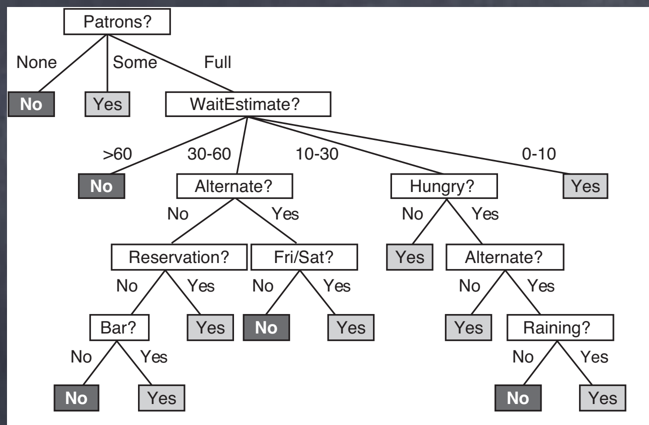


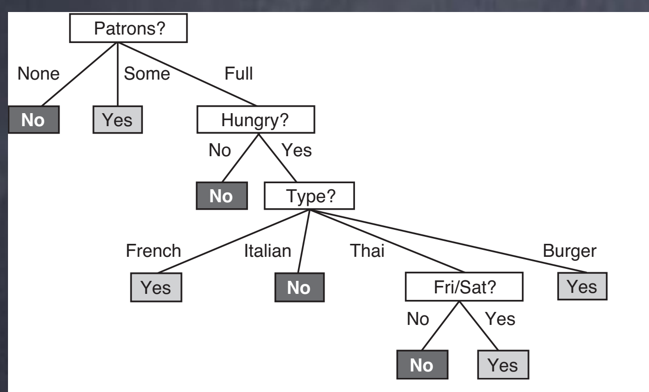
$$\mathbf{P}(R_0)$$

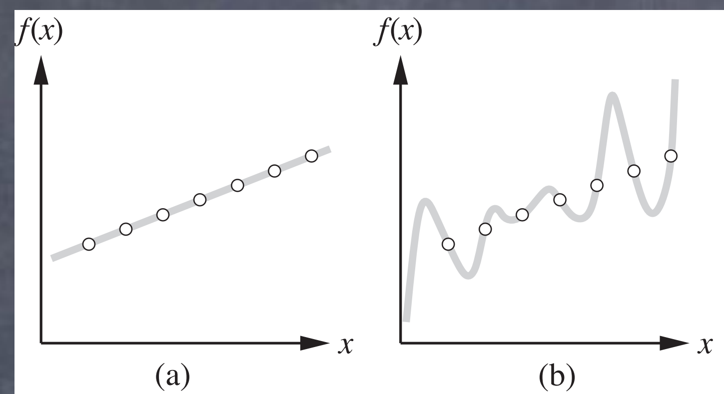
$$\mathbf{P}(R_t \mid R_{t\text{-}1})$$

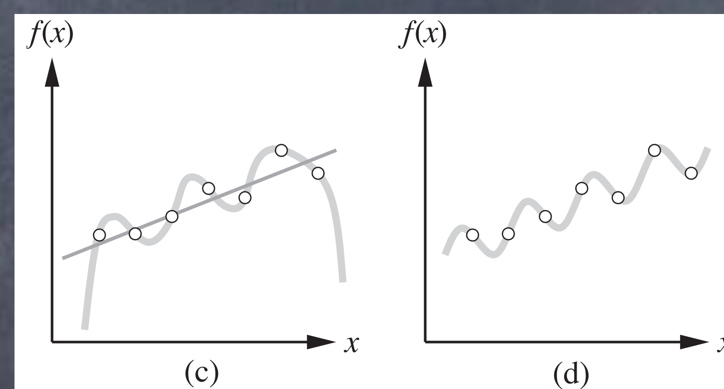
$$\mathbf{P}(U_t \mid R_t)$$

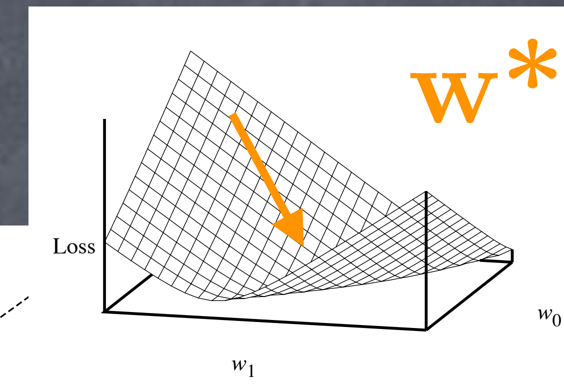
Learning

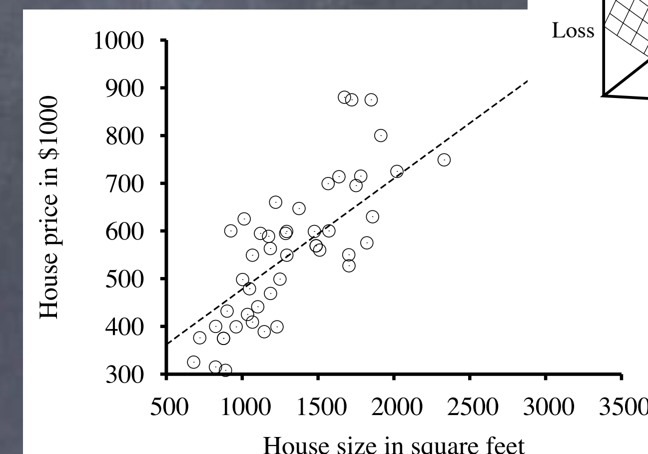




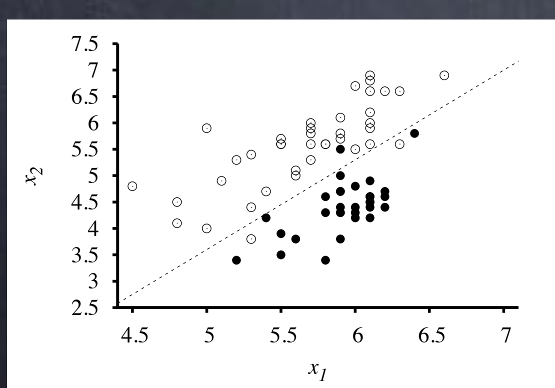


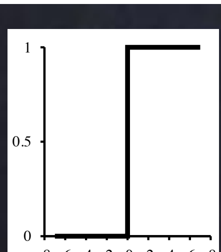


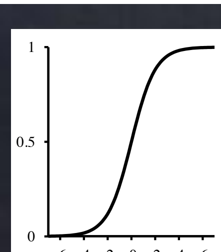


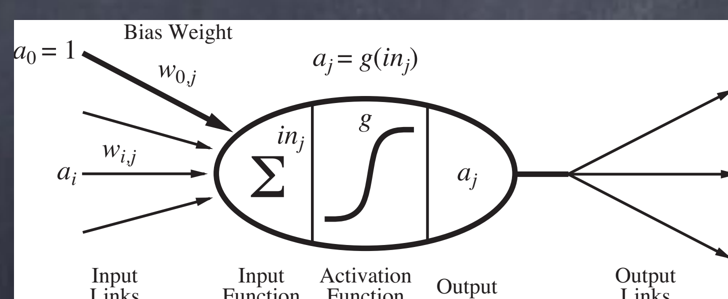


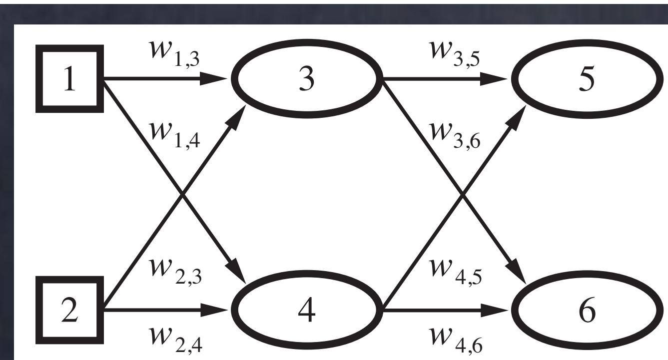
$$\underset{\Theta}{\operatorname{argmax}}\, P(\mathbf{d} \mid h_\Theta)$$



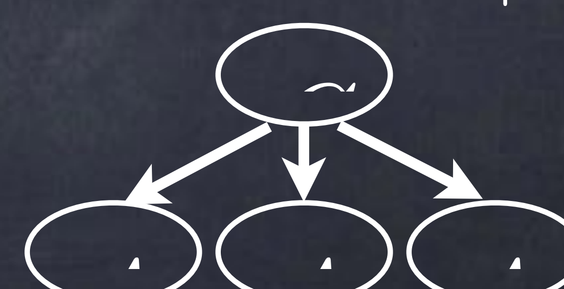








$$\Theta = \frac{c}{c+l} = \frac{c}{N}$$



Intro to AI

- Searching HUGE spaces for solutions to problems
- Representation matters!
  - Search through space of proofs
- Uncertainty: Can be done but...
  - Independence assumptions are the key
- Learning: Search for parameters

For Next Time

Unit 4 Exam

No class Tue 2 May

Final Exam: Thu 11 May 1600

Douglass Ballroom — BRING ID