Machine Learning 10-601

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

February 23, 2015

Today:

- Graphical models
- Bayes Nets:
  - Representing distributions
  - Conditional independencies
  - Simple inference
  - Simple learning

Readings:

- Bishop chapter 8, through 8.2
- Mitchell chapter 6

Dynamic BN   Time series ← RNN

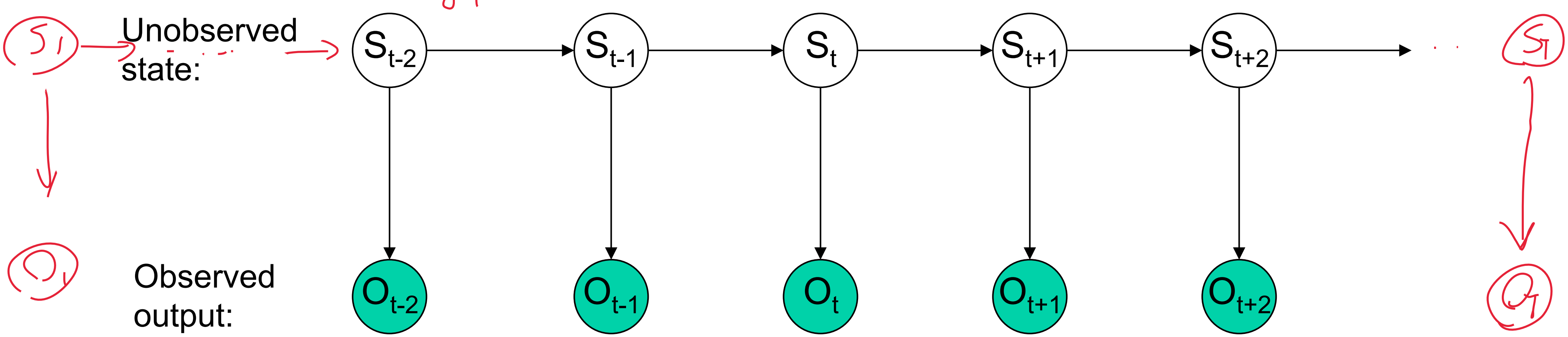
$A \perp\!\!\!\perp B \mid C \iff p(A,B \mid C) = p(A \mid C)\, p(B \mid C)$

Bayes Network for a Hidden Markov Model (HMM)

Implies the future is conditionally independent of the past, given the present

$S_t \perp\!\!\!\perp \{S_{t-2}, S_{t-3}, \dots, S_1\} \mid S_{t-1}$

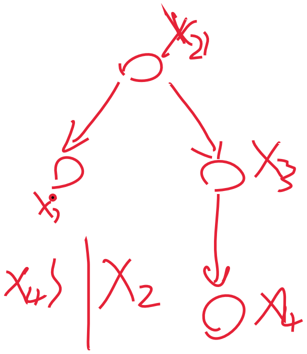
$P(X) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$



$$P(S_{t-2}, O_{t-2}, S_{t-1}, \dots, O_{t+2}) =$$

HMM: $p(S_1, O_1, \dots, S_T, O_T) = p(S_1)\, p(O_1 \mid S_1) \cdot \prod_{t=2}^{T} p(S_t \mid S_{t-1}) \cdot p(O_t \mid S_t)$

Conditional Independence, Revisited



$X_1 \perp\!\!\!\perp \{X_3, X_4\} \mid X_2$

- We said:
  - Each node is conditionally independent of its non-descendents, given its immediate parents.

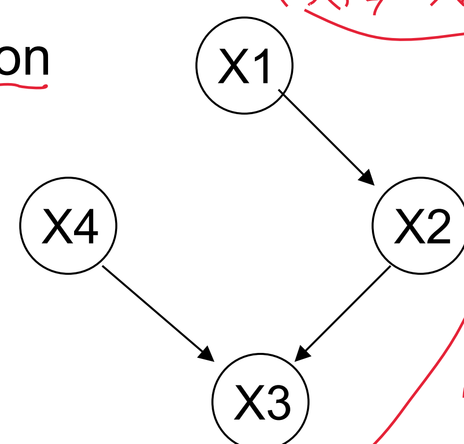
- Does this rule give us all of the conditional independence relations implied by the Bayes network?
  - No!
  - E.g., X1 and X4 are conditionally indep given {X2, X3}
  - But X1 and X4 not conditionally indep given X3
  - For this, we need to understand D-separation

$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$

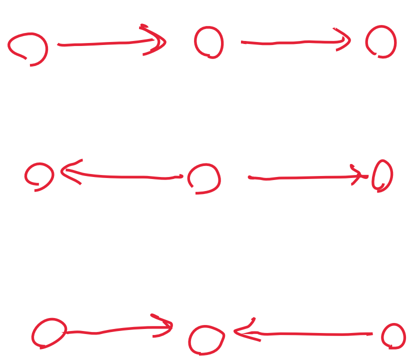
$X_1 \not\!\perp\!\!\!\perp X_4 \mid X_3$

D-separation

Quiz.



D.S



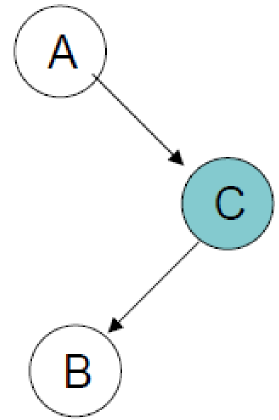
Head to Tail

Tail to Tail

Head to Head.

$$p(X,Y|Z) = p(X|Z)\,p(Y|Z)$$
$$X \perp\!\!\!\perp Y \mid Z$$

Easy Network 1: Head to Tail



prove A cond indep of B given C?

ie., p(a,b|c) = p(a|c) p(b|c)

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)\,p(c|a)\,p(b|c)}{p(c)}$$

$$= \frac{p(a,c)}{p(c)} = p(a|c)$$

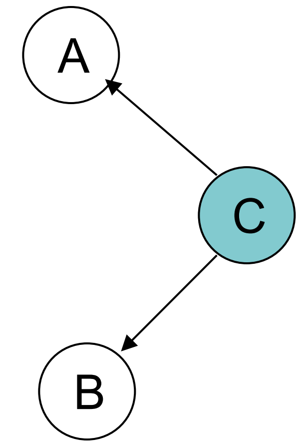
$$p(a,b) \overset{?}{\neq} p(a)\,p(b)$$

① $A \perp\!\!\!\perp B \mid C$

② $\cancel{A \not\perp\!\!\!\perp B}$

let's use p(a,b) as shorthand for p(A=a, B=b)

Easy Network 2: Tail to Tail



prove A cond indep of B given C? ie., p(a,b|c) = p(a|c) p(b|c) ✓

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{\cancel{p(c)}\, p(a|c)\, p(b|c)}{\cancel{p(c)}}$$

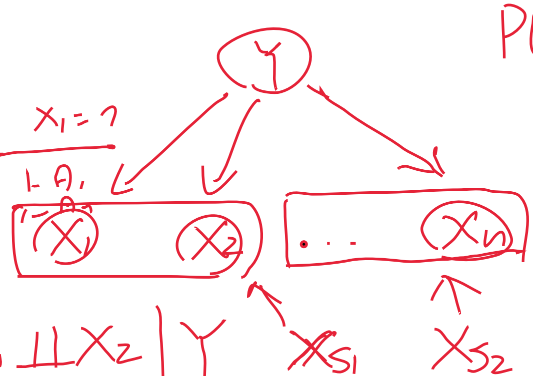
$$p(a,b) \neq p(a)\cdot p(b)$$

$$\boxed{\begin{array}{l} A \perp\!\!\!\perp B \mid C \\ A \not\perp\!\!\!\perp B \end{array}}$$

Naive Bayes

CPT

	$X_1=1$	$X_1=0$
$Y=1$	θ_1	$1-\theta_1$
$Y=0$	θ_n	



$$P(X,Y) = \underline{p(X|Y)}\, p(Y)$$

$$\Downarrow$$

$$= \prod_{i=1}^{n} p(X_i|Y)\, p(Y)$$

$X_1 \perp\!\!\!\perp X_2 \mid Y$

$X_i \perp\!\!\!\perp X_j \mid Y \quad (\forall\, i \neq j)$

$p(X_i, X_j|Y) = p(X_i|Y)\, p(X_j|Y)$

$$\boxed{X_{S_1} \perp\!\!\!\perp X_{S_2} \mid Y}$$

$$P(X|Y) \stackrel{?}{=} P(X_{S_1}|Y)\, P(X_{S_2}|Y)$$

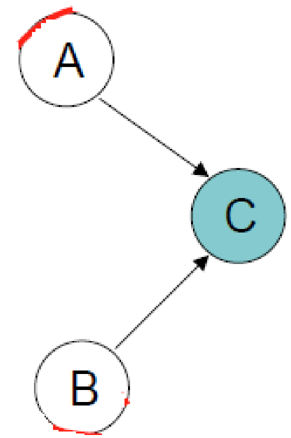
$$\Downarrow$$

$$\underbrace{\left(\prod_{i=1}^{2} p(X_i|Y)\right)}_{P(X_{S_1}|Y)} \underbrace{\left(\prod_{j=3}^{n} p(X_j|Y)\right)}_{P(X_{S_2}|Y)}$$

let's use p(a,b) as shorthand for p(A=a, B=b)

Easy Network 3: Head to Head

prove A cond indep of B given C? ie., $p(a,b|c) = p(a|c)\,p(b|c)$



$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \boxed{\frac{p(a)\,p(b)\,p(c|a,b)}{p(c)}}$$

$$A \not\perp\!\!\!\perp B \mid C$$

✓ $A \perp\!\!\!\perp B$ $p(a,b) = p(a)\,p(b)$

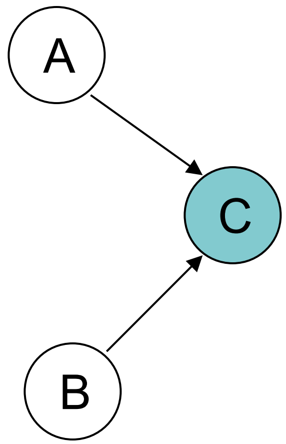
$$p(A=a, B=b) = p(A=a, B=b, C=1) + p(A=a, B=b, C=0)$$

$$= \underline{p(A=a)\,p(B=b)} \cdot p(C=1|A=a, B=b) + \underline{p(A=a)\,p(B=b)}\,p(C=0|A=a, B=b)$$

$$= p(A=a)\,p(B=b)$$

let's use p(a,b) as shorthand for p(A=a, B=b)

Easy Network 3: Head to Head



prove A cond indep of B given C?   NO!

Summary:

- p(a,b)=p(a)p(b) $A \perp\!\!\!\perp B$
- p(a,b|c) NotEqual p(a|c)p(b|c) $A \not\!\perp\!\!\!\perp B \mid C$

Explaining away.

e.g.,

- A=earthquake
- B=breakIn
- C=motionAlarm

X and Y are conditionally independent given Z, **<u>if and only if</u>** X and Y are D-separated by Z.

[Bishop, 8.2.2]

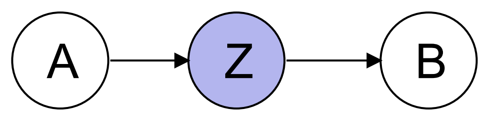
Suppose we have three sets of random variables: X, Y and Z

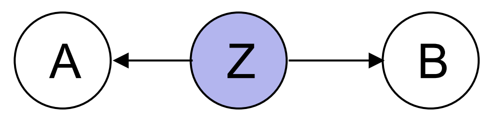
X and Y are **<u>D-separated</u>** by Z (and therefore conditionally indep, given Z) iff every path from every variable in X to every variable in Y is **<u>blocked</u>**

↳ < H-T, T-T
H-H

A path from variable X to variable Y is **blocked** if it includes a node in Z such that either

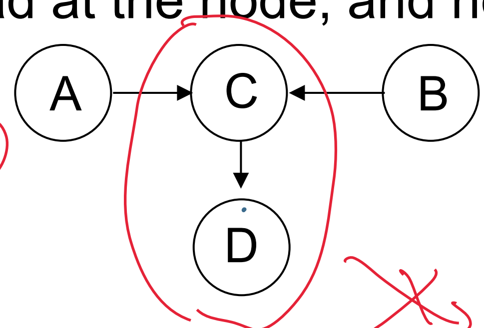
$A \perp\!\!\!\perp B \mid Z$





1. arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

2. or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z



$A \not\perp\!\!\!\perp B \mid C$

$A \not\perp\!\!\!\perp B \mid D$

Two kinds of pathes (A — B)

① H-T, T-T → Z

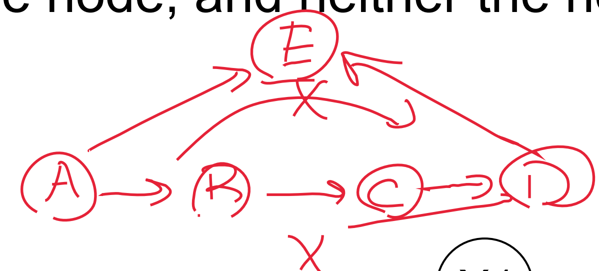
② H-H ↛ Z

X and Y are **D-separated** by Z (and therefore conditionally indep, given Z) iff every path from every variable in X to every variable in Y is **blocked**

A path from variable A to variable B is **blocked** if it includes a node such that either

1.arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

2.or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z



$A \perp\!\!\!\perp D \mid B$

$A \perp\!\!\!\perp D \mid C$

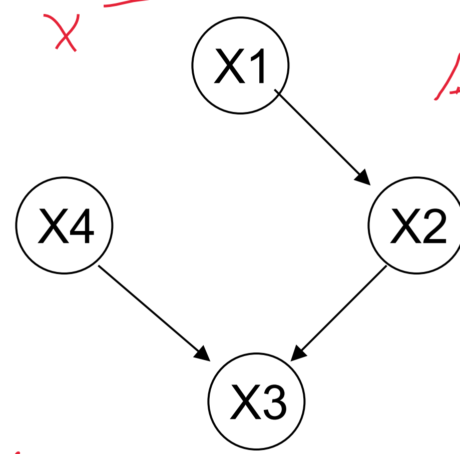
$A \not\perp\!\!\!\perp D \mid E, B$

$X_1 \perp\!\!\!\perp X_3 \mid X_2$

X1 indep of X3 given X2?

X3 indep of X1 given X2?

X4 indep of X1 given X2? $X_1 \perp\!\!\!\perp X_4 \mid X_2$



$X_1 \perp\!\!\!\perp X_4$

D-separation

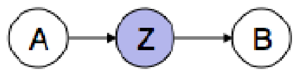
① H-T, T-T $\rightarrow$ Z

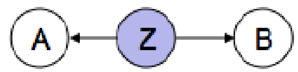
② H-H $\not\rightarrow$ Z

X and Y are **<u>D-separated</u>** by Z (and therefore conditionally indep, given Z) iff every path from any variable in X to any variable in Y is **<u>blocked</u>** by Z

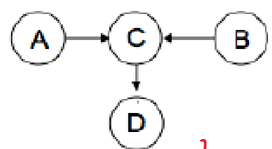
A path from variable A to variable B is **blocked** by Z if it includes a node such that either

1.arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z





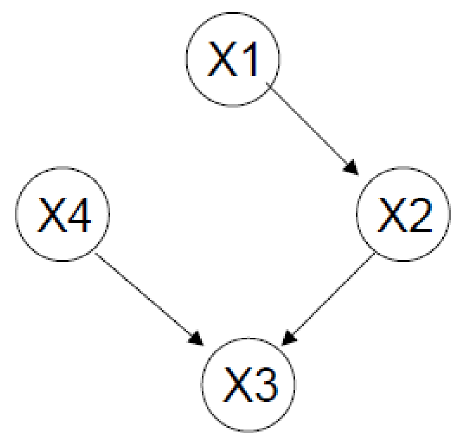
2.the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z



X4 indep of X1 given X3? $X_4 \not\perp\!\!\!\perp X_1 \mid X_3$

X4 indep of X1 given {X3, X2}?

X4 indep of X1 given {}? ✓



X and Y are **<u>D-separated</u>** by Z (and therefore conditionally indep, given Z) iff every path from any variable in X to any variable in Y is **<u>blocked</u>**

A path from variable A to variable B is **blocked** if it includes a node such that either

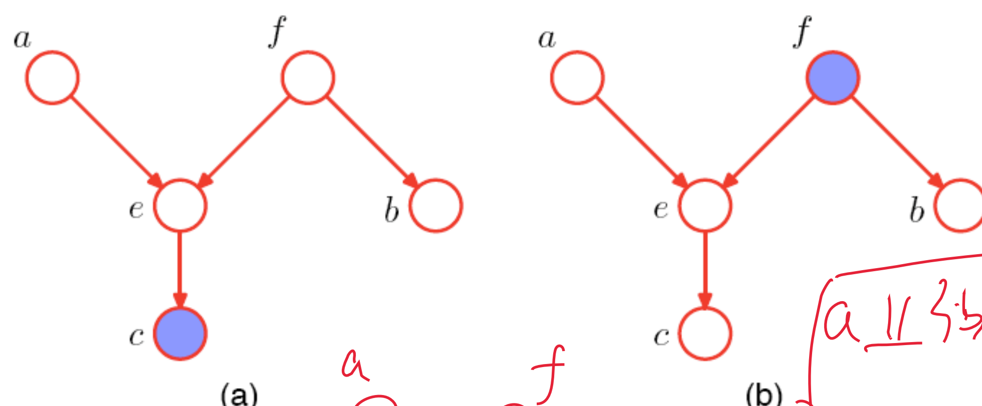
1. arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

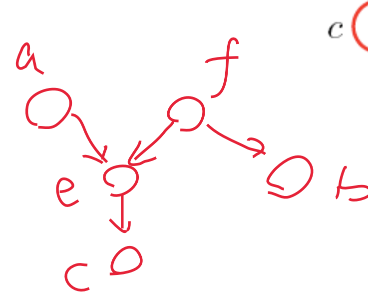
2. or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z

a indep of b given c? ✗

a indep of b given f ? ✓

$a \perp\!\!\!\perp b \mid f$





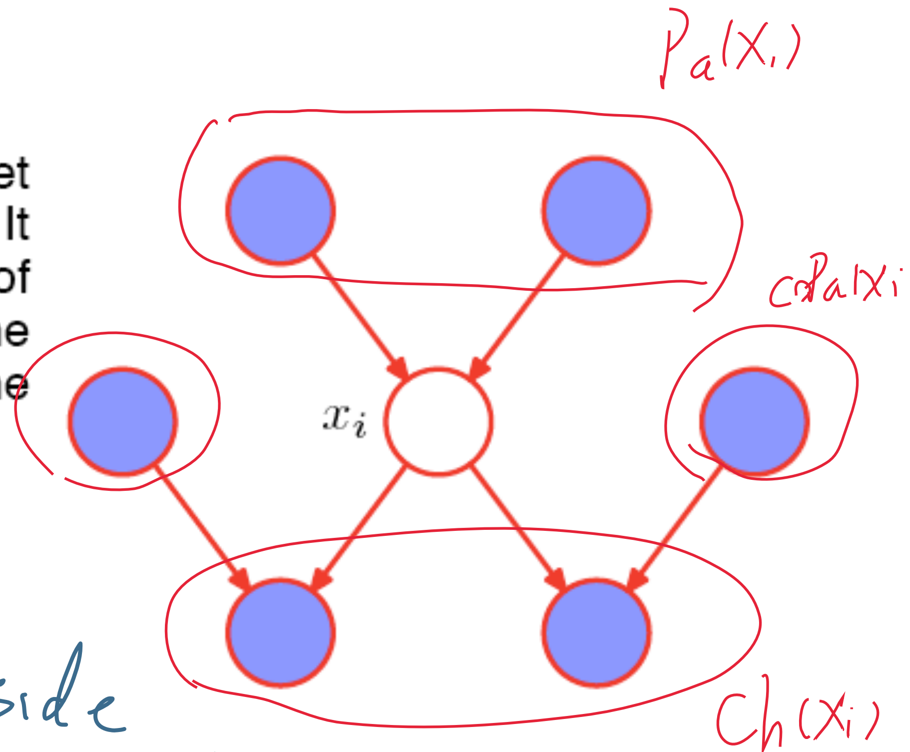
$a \perp\!\!\!\perp \{b, f\} \mid \phi$

$a \perp\!\!\!\perp b \mid \phi$

$a \perp\!\!\!\perp f \mid \phi$

Markov Blanket

The Markov blanket of a node $\mathbf{x}_i$ comprises the set of parents, children and co-parents of the node. It has the property that the conditional distribution of $\mathbf{x}_i$, conditioned on all the remaining variables in the graph, is dependent only on the variables in the Markov blanket.



co-parent = other side of X_i's colliders

$$P(X_i \mid X_{\{j \neq i\}})$$

$$= P(X_i \mid X_{\{j \in \overline{MB_i}\}}, X_{\{k \in MB_i\}})$$

D-separation

$$= P(X_i \mid X_{\{k \in MB_i\}}) \qquad X_i \perp\!\!\!\perp X_{\{j \in \overline{MB_i}\}} \mid X_{\{k \in MB_i\}}$$

from [Bishop, 8.2]

What You Should Know

- Bayes nets are convenient representation for encoding dependencies / conditional independence
- BN = Graph plus parameters of CPD's (CPT)
  - Defines joint distribution over variables
  - Can calculate everything else from that
  - Though inference may be intractable
- Reading conditional independence relations from the graph
  - Each node is cond indep of non-descendents, given only its parents
  - X and Y are conditionally independent given Z if Z D-separates every path connecting X to Y
  - Marginal independence : special case where Z={}

H-T
T-T
H-H

$O(\log n) < O(n) < O(n^2) < O(e^n) < O(n!)$ $n \to \infty$

Inference in Bayes Nets

- In general, intractable (NP-~~complete~~ hard)
- For certain cases, tractable

exact

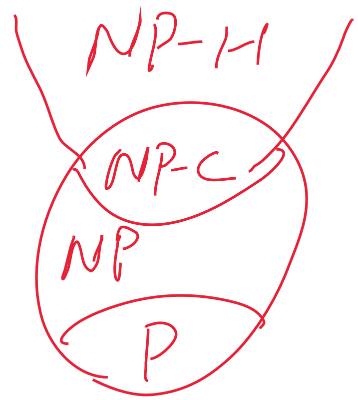
  - Assigning probability to fully observed set of variables $p(a,b,c) = \_\_\ \_\_\ \_\_$
  - Or if just one variable unobserved $p(a,b) = \sum_c p(a, b, C=c)$
  - Or for singly connected graphs (ie., no undirected loops)
    - Belief propagation sum-prod / max-sum
- Sometimes use Monte Carlo methods

approximate

  - Generate many samples according to the Bayes Net distribution, then count up the results
- Variational methods for tractable approximate solutions

完全 困难

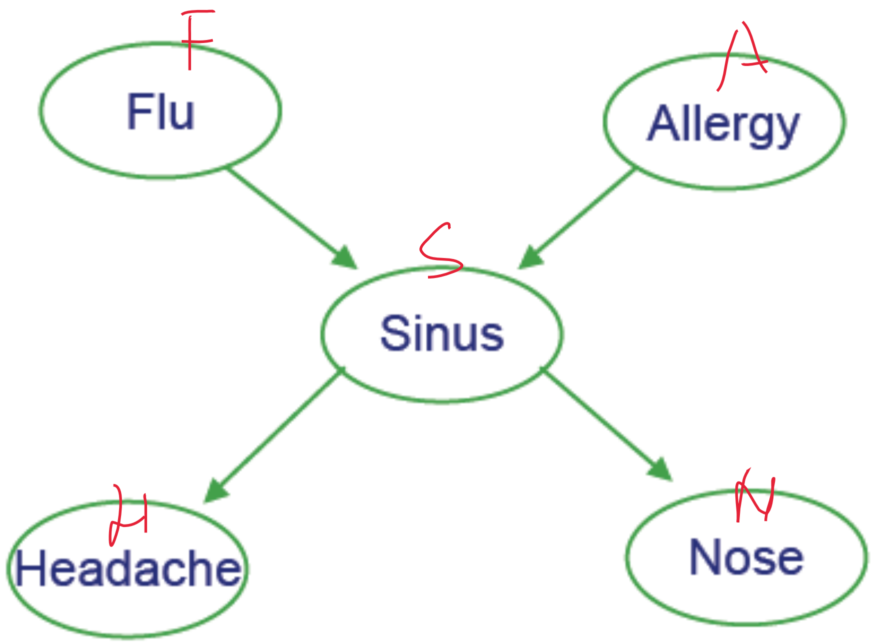
P NP NP-C NP-H



$P \neq NP$

Example

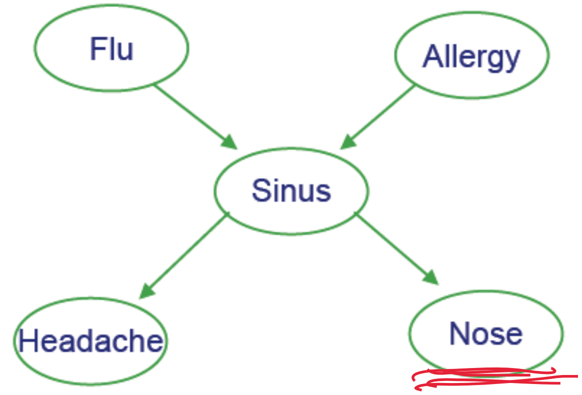
- Bird flu and Allegies both cause Sinus problems
- Sinus problems cause Headaches and runny Nose



BN < DAG
CPD

$P(N \mid S)$	$N=1$	$N=0$
$S=1$	.	-
$S=0$	-	-

Prob. of joint assignment: easy



fully-observed

- Suppose we are interested in joint assignment <F=f,A=a,S=s,H=h,N=n>

What is P(f,a,s,h,n)?

Boolean.

$(n-1)$: Multiplication

$$p(f,a,s,h,n) = p(f)\,p(a)\,p(s|f,a)\,\cancel{p(h|s)}\,p(n|s)$$

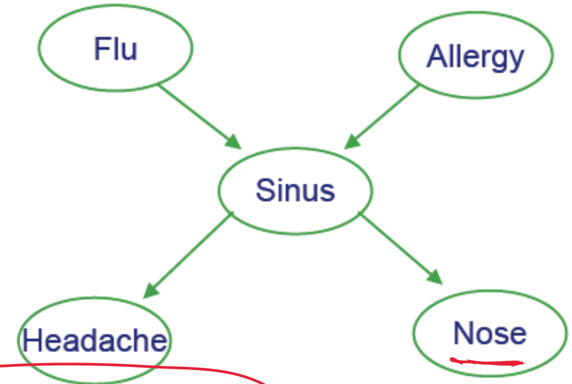
$$= \_\_\_$$

$$p(f,a,s,h) = \sum_{n} p(f,a,s,h,N=n)$$

$O(n)$

let's use p(a,b) as shorthand for p(A=a, B=b)

Prob. of marginals: not so easy



- How do we calculate P(N=n) ?

summing JD

$$\boxed{p(N=n) = \sum_{f,a,s,h} p(F=f, A=a, S=s, H=h, N=n)}$$

$2^4 = 16$

$p(F=0, A=0, S=0, H=0, N=n)$ $(n-1)$

$= p(f)\,p(a)\,p(s|f,a)\,p(h|s)\cdot p(N=n|s)$

$\vdots$

$p(F=1, A=1, S=1, H=1, N=n)$

n var:

summation:

2^n

multiplication:

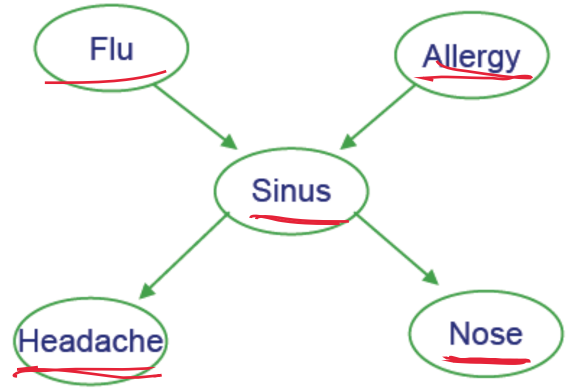
$2^n \cdot (n-1)$

$O(n \cdot 2^n) > O(2^n)$

NP hard

let's use p(a,b) as shorthand for p(A=a, B=b)

Generating a sample from joint distribution: easy



How can we generate random samples drawn according to P(F,A,S,H,N)?

Hint: random sample of F according to $P(F=1) = \theta_{F=1}$: $\in [0,1]$

- draw a value of r uniformly from [0,1]
- if $r<\theta$ then output F=1, else F=0

0 $\theta = 0.7$ 1

$(f, a, s, h, n) \in \mathbb{B}^5$

(F=0, A≠0)

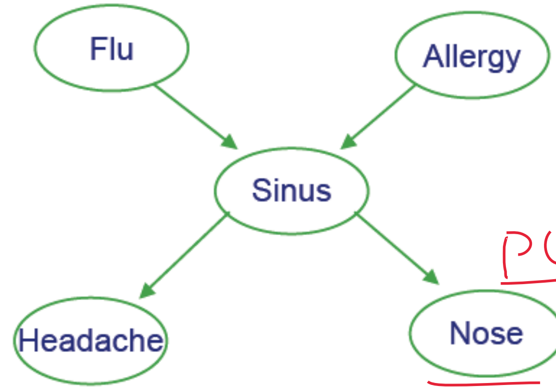
$P(S \mid F,A)$	S=1	S=0
F=0, A=0	θ_{00}	$1-\theta_{00}$
F=0, A=1	θ_{01}	
F=1, A=0	θ_{10}	
F=1, A=1	θ_{11}	

let's use p(a,b) as shorthand for p(A=a, B=b)

D

MLE MAP.

Generating a sample from joint distribution: easy



$P(N|S)$

How can we generate random samples drawn according to P(F,A,S,H,N)?

$P(N=n)$

$= \theta^{N}(1-\theta)^{1-N}$

Hint: random sample of F according to $P(F=1) = \theta_{F=1}$:

- draw a value of r uniformly from [0,1]
- if $r<\theta$ then output F=1, else F=0

x_i

$D = \{(f_i, a_i, s_i, h_i, n_i)\}_{i=1}^{N}$

$L(\theta) = P(D|\theta) = \prod_{i=1}^{N} P(N_i|\theta)$

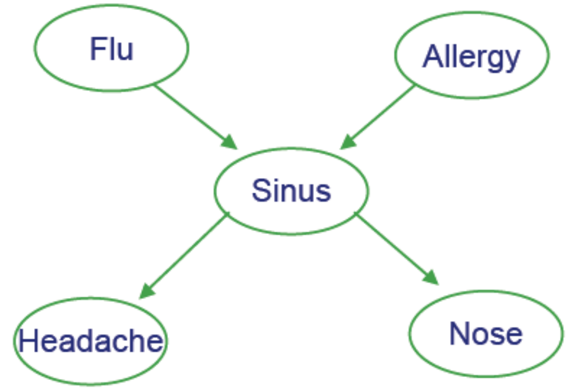
Solution:

- draw a random value f for F, using its CPD
- then draw values for A, for S|A,F, for H|S, for N|S

$\hat{\theta} = \frac{|D_{N=n}|}{|D|}$

MLE

Generating a sample from joint distribution: easy



Note we can estimate marginals

like P(N=n) by generating many samples

from joint distribution, then count the fraction of samples for which N=n

Similarly, for anything else we care about

$P(F=1|H=1, N=0) = \frac{|D_{F=1,H=1,N=0}|}{|D_{H=1,N=0}|}$

MLE

→ weak but general method for estimating any probability term…

Inference in Bayes Nets

H-T H-H

$A \rightarrow B \rightarrow C \leftarrow D$

$A \perp\!\!\!\perp D \mid B$

$A \perp\!\!\!\perp D \mid \phi$

$A \not\perp\!\!\!\perp D \mid C$

- In general, intractable (NP-complete)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
  - Or for singly connected graphs (ie., no undirected loops)
    - Variable elimination
    - Belief propagation
- Often use Monte Carlo methods
  - e.g., Generate many samples according to the Bayes Net distribution, then count up the results
  - Gibbs sampling
- Variational methods for tractable approximate solutions

see Graphical Models course 10-708