Readings:
K&F: 3.1, 3.2, 3.3

BN Semantics 1

Graphical Models – 10708
Carlos Guestrin
Carnegie Mellon University
September 15th, 2008

Let's start on BNs…

- Consider $P(X_i)$
  - Assign probability to each $x_i \in Val(X_i)$
  - Independent parameters $|Val(X_i)| = k$
    $k-1$
- Consider $P(X_1,\dots,X_n)$
  - How many independent parameters if $|Val(X_i)|=k$?
    $k^n - 1$
    same thing w. fewer params } BN

What if variables are independent?

- What if variables are independent?
  - $(X_i \perp X_j)$, $\forall$ i,j
  - Not enough!!! (See homework 1 ☺)
  - Must assume that $(\mathbf{X} \perp \mathbf{Y})$, $\forall$ **X**,**Y** subsets of $\{X_1,\ldots,X_n\}$

$X_1 X_3 \perp X_7 X_{14}$

$X_1 X_{14} \perp X_3 X_2 X_5$

- Can write
  - $P(X_1,\ldots,X_n) = \prod_{i=1\ldots n} P(X_i)$

BN w. no edges

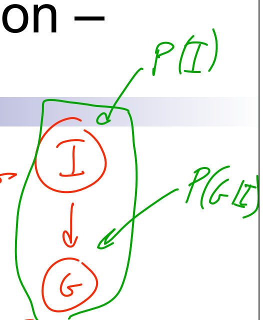
X_1 X_2 X_3 . . . X_n

- How many independent parameters now?

$n.(k-1)$

Conditional parameterization – two nodes

- Grade is determined by Intelligence



$P(I) =$

VH	H
.85	.15

$P(G|I) =$

G \ I	VH	H
A	.9	.5
B	.1	.5

$P(I{=}VH, G{=}B)$

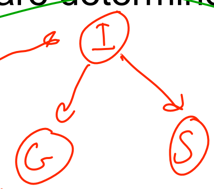
$= P(I{=}VH)\, P(G{=}B \mid I{=}VH)$

$= .85 \times .1$

$= 0.085$

Conditional parameterization – three nodes

- Grade and SAT score are determined by Intelligence
- $(G \perp S \mid I)$



$P(S \mid G,I) = P(S \mid I)$

$P(I)$ $\quad$ $P(G \mid I)$ $\quad$ $P(S \mid I)$

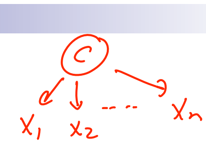
$P(I,G,S) = P(I) \cdot P(G \mid I)\, P(S \mid I)$ why??

Chain rule $P(I,G,S) = P(I)\, P(G \mid I) \cdot P(S \mid G,I)$

$= P(I)\, P(G \mid I)\, P(S \mid I)$ same because of $(G \perp S \mid I)$

The naïve Bayes model – Your first real Bayes Net

- Class variable: C
- Evidence variables: $X_1,\ldots,X_n$
- assume that $(\mathbf{X} \perp \mathbf{Y} \mid C)$, $\forall$ $\mathbf{X},\mathbf{Y}$ subsets of $\{X_1,\ldots,X_n\}$



$P(C,X_1,\ldots X_n) = P(C) \prod_i P(X_i \mid C)$ ← why??

$P(C,X_1 \ldots X_n) = P(C)\, P(X_1 \mid C) \cdot P(X_2 \mid C,X_1) \ldots P(X_n \mid C,X_1 \ldots X_{n-1})$ ← no assumptions

$P(X_2 \mid C,X_1) = P(X_2 \mid C)$

$X_1 \perp X_2 \mid C$

$P(X_n \mid C,X_1 \ldots X_{n-1}) = P(X_n \mid C)$

$X_n \perp X_1 \ldots X_{n-1} \mid C$

becomes Naïve Bayes model

What you need to know (From last class)

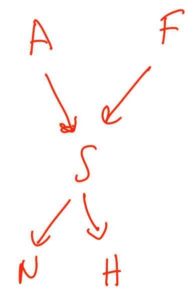
- Basic definitions of probabilities
- Independence
- Conditional independence
- The chain rule
- Bayes rule
- Naïve Bayes

This class

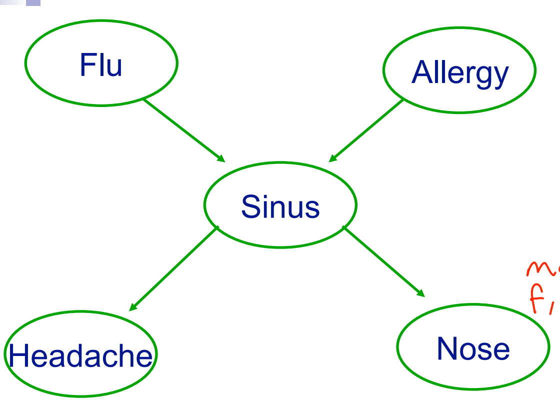
- We've heard of Bayes nets, we've played with Bayes nets, we've even used them in your research
- This class, we'll learn the semantics of BNs, relate them to independence assumptions encoded by the graph

Causal structure

- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches
- How are these connected?

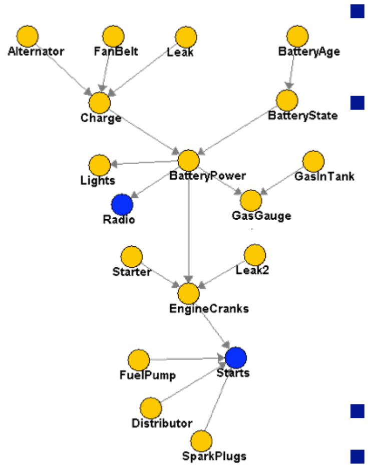


Possible queries



- Probabilistic Inference
  $H=t$, $N=f$
  $P(A=t \mid H=t, N=f)$
- Most probable explanation
  $\max_{f,a,s} P(F=f, A=a, S=s \mid H=t, N=f)$
- Active data collection
  ↳ what's next best test?
  variable to observe

Car starts BN

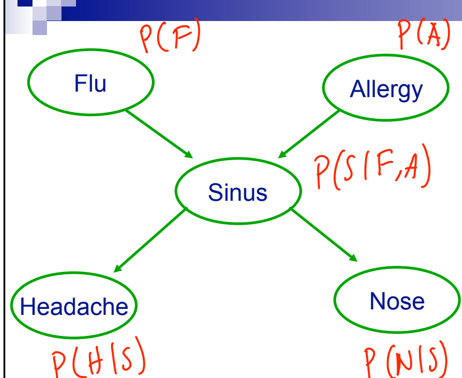


- 18 binary attributes
  - $P(A, F, L, \ldots \text{Spark})$
- Inference
  - P(BatteryAge|Starts=f)
  - marginalize all variables except for BatteryAge, Starts

$$= \underbrace{\sum_{Alt} \sum_{FanBelt} \sum_{\ldots}}_{2^{16}} \ldots P(A, F, L, \ldots)$$

- 2^{18} terms, why so fast?
  - Sparse structure!! :)
- Not impressed?
  - HailFinder BN – more than 3^{54} = 58149737003040059690390169 terms

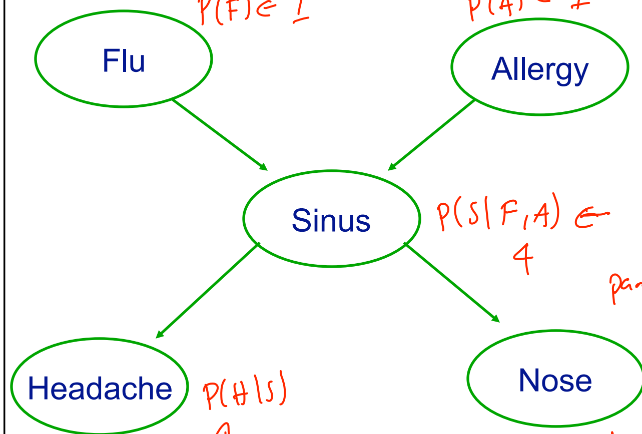
Factored joint distribution - Preview



$$P(F, A, S, H, N) = P(F) \cdot P(A) \cdot P(S|F,A) \cdot P(H|S) \cdot P(N|S)$$

Number of parameters

binary vars



$P(F) \leftarrow 1$

$P(A) \leftarrow 1$

$P(S|F,A) \leftarrow 4$

$P(H|S) \leftarrow 2$

$P(N|S) \leftarrow 2$

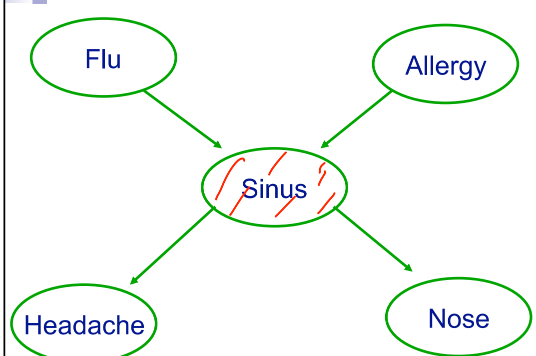
$P(F,A,S,H,N)$ explicitly $2^5 - 1 = 31$

S \ FA	tt	tf	ft	ff
t	.9	.7	.8	.2
f	.1	.3	.2	.8

params $1 + 1 + 1 + 1 = 4$

with BN: 10

Key: Independence assumptions



$\neg\, H \perp N$

$H \perp N \mid S$

$\neg\, A \perp N$

$A \perp N \mid S$

Knowing sinus separates the symptom variables from each other

(Marginal) Independence

- Flu and Allergy are (marginally) independent

$F \perp A$

$P(F,A) = P(F).P(A)$

Flu = t	0.1
Flu = f	0.9

Allergy = t	.3
Allergy = f	.7

	Flu = t	Flu = f
Allergy = t	$.1 \times .3 = 0.03$	$.3 \times .9$
Allergy = f	$.1 \times .7$	$.9 \times .7$

- More Generally:

$\forall$ subsets of $X_1 \ldots X_n$

$X \perp Y$ $\quad X \subseteq X_1 \ldots X_n$, $Y \subset X_1 \ldots X_n$

$P(X_1 \ldots X_n) = \prod_i P(X_i)$

Conditional independence

- Flu and Headache are not (marginally) independent

$P(F|H) \neq P(F)$

$F \to S \to H$

$\neg F \perp H$

$P(F,H) \neq P(F).P(H)$

- Flu and Headache are independent given Sinus infection

$F \perp H | S$

$P(F,H|S) = P(F|S)\ P(H|S)$

$P(F|H,S) = P(F|S)$

- More Generally:

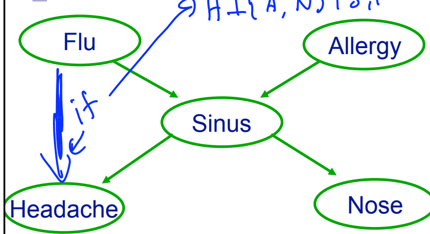
$X_1 \perp X_2 \ldots X_n | C \iff P(X_1, X_2 \ldots X_n | C) = P(X_1|C).P(X_2 \ldots X_n|C)$ OR $P(X_1|C, X_2 \ldots X_n) = P(X_1|C)$

The independence assumption

Before new edge $H \perp \{A, F, N\} \mid S$

$H \perp \{A, N\} \mid S, F$

if



Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents
($X_i \perp NonDescendants_{X_i} \mid Pa_{X_i}$)

Flu: $Pa_{Flu} = \emptyset$
$NonDescendants_{Flu} = \{A\}$
$F \perp A$

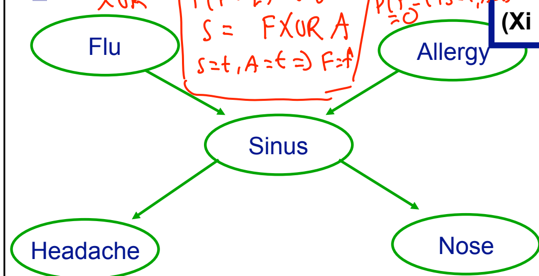
Nose: $Pa_{Nose} = \{S\}$
$NonDescendants_{Nose} = \{F, A, H\}$
$N \perp \{F, A, H\} \mid S$

Sinus: $Pa_{Sinus} = \{F, A\}$
$NonDescendants_{Sinus} = \emptyset$
NO ASSUMPTIONS ABOUT $S \perp ?? \mid F, A$

Explaining away

Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents
($X_i \perp NonDescendants_{X_i} \mid Pa_{X_i}$)

XOR
$P(A=t) = 0.5$
$P(F=t) = 0.5$
$S = F \text{ XOR } A$
$S=t, A=t \Rightarrow F=f$
$P(F=t \mid S=t, A=t) = 0$



$P(F=t) = .2$
$P(F=t \mid S=t) = .5$
Now $S=t$, $A=t$
$P(F=t \mid S=t, A=t) = .3$

$F \perp A$
S is not a parent!
$\neg\ F \perp A \mid S$
not implied by Local Markov assumption

$P(F=t) \leq \underbrace{P(F=t \mid S=t, A=t) \leq P(F=t \mid S=t)}$
for this example
could be $P(F=t \mid S=t, A=t) \geq P(F=t \mid S=t)$
depends on $P(S \mid F, A)$