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 $P(X_1,\ldots,X_n)$
  - How many independent parameters if $|Val(X_i)|=k$?
    $k^n - 1$
    Same thing w. fewer params } BN

What if variables are independent?

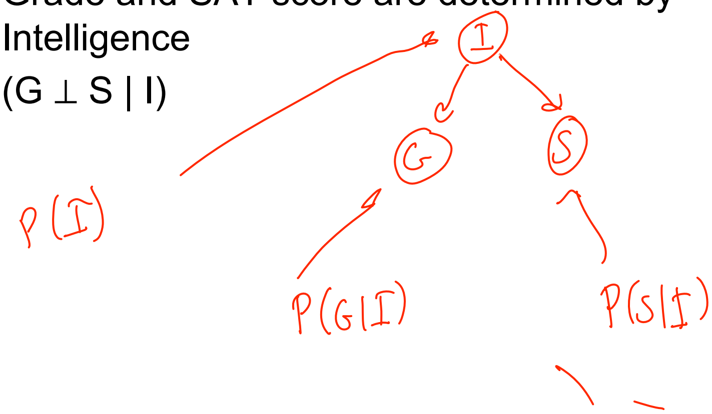
- What if variables are independent?
  - $(X_i \perp X_j), \forall i,j$
  - Not enough!!! (See homework 1 ☺)
  - Must assume that $(\mathbf{X} \perp \mathbf{Y}), \forall\, \mathbf{X},\mathbf{Y}$ subsets of $\{X_1,\ldots,X_n\}$
- Can write
  - $P(X_1,\ldots,X_n) = \prod_{i=1\ldots n} P(X_i)$
- How many independent parameters now?

Conditional parameterization – two nodes

- Grade is determined by Intelligence

Conditional parameterization – three nodes

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



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 $\{X_1,\ldots,X_n\}$

What you need to know (From last class)

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

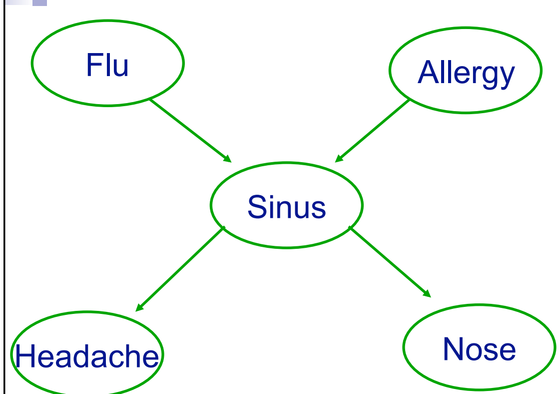
This class

- 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

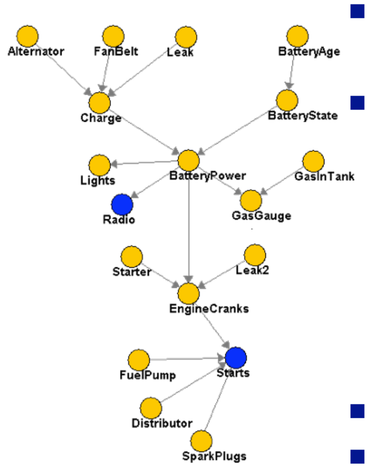
- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches
- How are these connected?

Possible queries



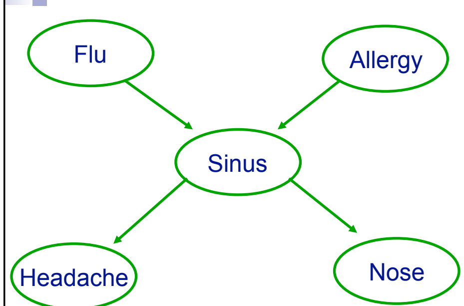
- Inference
- Most probable explanation
- Active data collection

Car starts BN

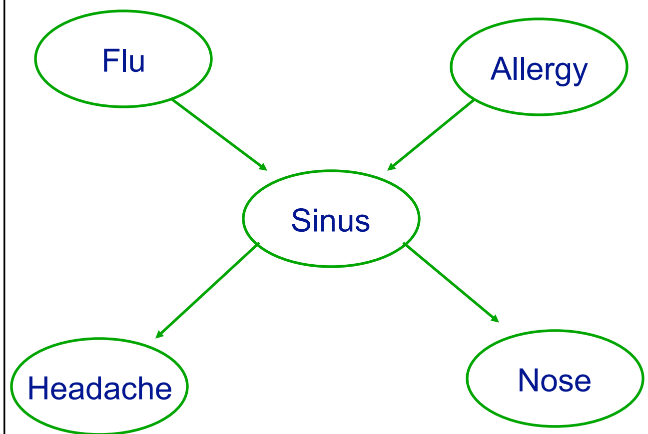


- 18 binary attributes
- Inference
  - P(BatteryAge|Starts=f)
- 2^{18} terms, why so fast?
- Not impressed?
  - HailFinder BN – more than 3^{54} = 58149737003040059690390169 terms

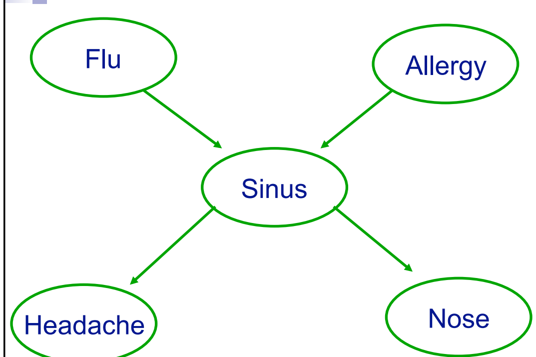
Factored joint distribution - Preview



Number of parameters



Key: Independence assumptions



Knowing sinus separates the symptom variables from each other

(Marginal) Independence

- Flu and Allergy are (marginally) independent

Flu = t	
Flu = f	

Allergy = t	
Allergy = f	

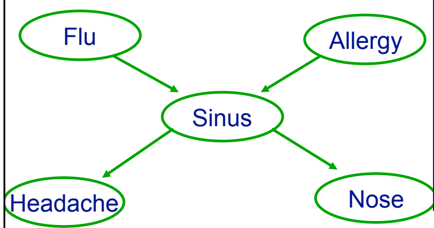
	Flu = t	Flu = f
Allergy = t		
Allergy = f		

- More Generally:

Conditional independence

- Flu and Headache are not (marginally) independent
- Flu and Headache are independent given Sinus infection
- More Generally:

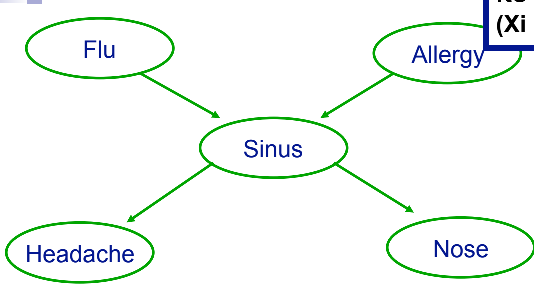
The independence assumption



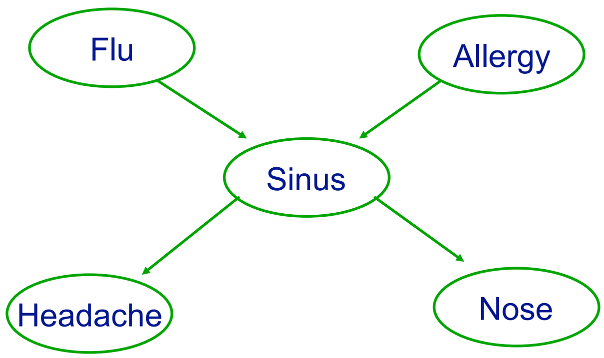
Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents
($X_i \perp NonDescendants_{Xi} \mid Pa_{Xi}$)

Explaining away

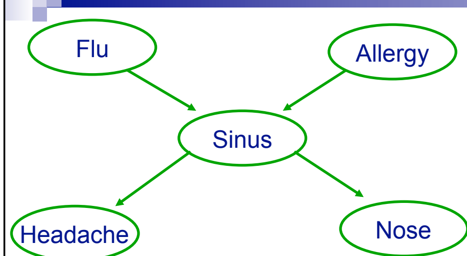
Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents
($Xi \perp NonDescendants_{Xi} \mid Pa_{Xi}$)



What about probabilities? Conditional probability tables (CPTs)



Joint distribution



Why can we decompose? Local Markov Assumption!

A general Bayes net

- Set of random variables
- Directed acyclic graph
- CPTs
- Joint distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

- **Local Markov Assumption:**
  - A variable X is independent of its non-descendants given its parents and only its parents – **(Xi ⊥ NonDescendantsXi | PaXi)**

Announcements

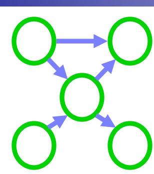
- Homework 1:
  - Out wednesday
  - Due in 2 weeks – **beginning of class!**
  - It's hard – start early, ask questions
- Collaboration policy
  - OK to discuss in groups
  - Tell us on your paper who you talked with
  - Each person must write their **own unique paper**
  - No searching the web, papers, etc. for answers, we trust you want to learn
- Audit policy
  - No sitting in, official auditors only, see couse website
- Don't forget to register to the mailing list at:
  - https://mailman.srv.cs.cmu.edu/mailman/listinfo/10708-announce

Questions????

- What distributions can be represented by a BN?
- What BNs can represent a distribution?
- What are the independence assumptions encoded in a BN?
  - in addition to the local Markov assumption

Today: The Representation Theorem – Joint Distribution to BN

BN:



Encodes independence assumptions

If conditional independencies in BN are subset of conditional independencies in *P*

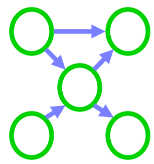
Obtain

Joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Today: The Representation Theorem – BN to Joint Distribution

BN:



Encodes independence assumptions

If joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain →

Then conditional independencies in BN are subset of conditional independencies in *P*

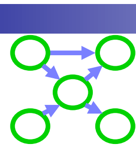
Let's start proving it for naïve Bayes – From joint distribution to BN

- Independence assumptions:
  - X_i independent given C
- Let's assume that *P* satisfies independencies must prove that *P* factorizes according to BN:
  - $P(C,X_1,\ldots,X_n) = P(C) \prod_i P(X_i|C)$
- Use chain rule!

Let's start proving it for naïve Bayes – From BN to joint distribution

- Let's assume that P factorizes according to the BN:
  - $P(C,X_1,\ldots,X_n) = P(C) \prod_i P(X_i|C)$
- Prove the independence assumptions:
  - X_i independent given C
  - Actually, $(\mathbf{X} \perp \mathbf{Y} \mid C)$, $\forall$ $\mathbf{X},\mathbf{Y}$ subsets of $\{X_1,\ldots,X_n\}$

Today: The Representation Theorem

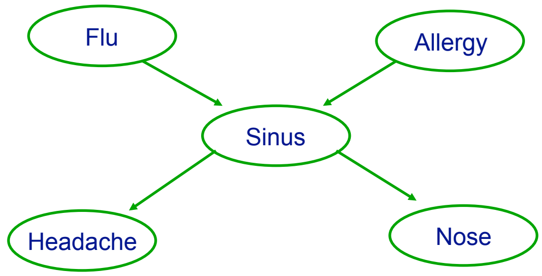
BN:  **Encodes independence assumptions**

If conditional independencies in BN are subset of conditional independencies in *P*	Obtain	**Joint probability distribution:** $P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$

If joint probability distribution: $P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$	Obtain	**Then conditional independencies in BN are subset of conditional independencies in *P***

Local Markov assumption & I-maps

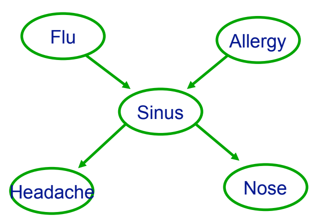
- Local independence assumptions in BN structure G:
- Independence assertions of *P*:
- BN structure G is an ***I-map*** (independence map) if:



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

Factorized distributions

- Given
  - Random vars $X_1,\ldots,X_n$
  - *P* distribution over vars
  - BN structure *G* over same vars
- *P* factorizes according to *G* if



$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

BN Representation Theorem – I-map to factorization

If conditional independencies in BN are subset of conditional independencies in *P*

Obtain →

Joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

G* is an I-map of *P

P* factorizes according to *G

BN Representation Theorem – I-map to factorization: **Proof**

G* is an I-map of *P

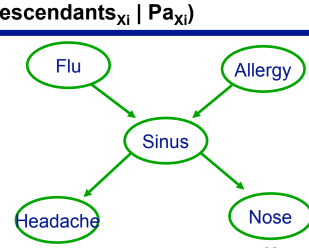
Obtain →

P* factorizes according to *G

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

ALL YOU NEED:

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



Defining a BN

- Given a set of variables and conditional independence assertions of P
- Choose an ordering on variables, e.g., X_1, …, X_n
- For i = 1 to n
  - Add X_i to the network
  - Define parents of X_i, $\mathbf{Pa}_{X_i}$, in graph as the minimal subset of $\{X_1,\dots,X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1, \dots, X_{i-1}\}$, given parents $\mathbf{Pa}_{Xi}$
  - Define/learn CPT – $P(X_i | \mathbf{Pa}_{Xi})$

BN Representation Theorem – Factorization to I-map

If joint probability distribution:

$$P(X_1,\dots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain

Then conditional independencies in BN are subset of conditional independencies in *P*

P* factorizes according to *G

G* is an I-map of *P

BN Representation Theorem – Factorization to I-map: **Proof**

If joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain →

Then conditional independencies in BN are subset of conditional independencies in ***P***

P* factorizes according to *G

G* is an I-map of *P

Homework 1!!!! ☺

The BN Representation Theorem

If conditional independencies in BN are subset of conditional independencies in ***P***

Obtain →

Joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Important because:
Every *P* has at least one BN structure *G*

If joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain →

Then conditional independencies in BN are subset of conditional independencies in ***P***

Important because:
Read independencies of *P* from BN structure *G*

Acknowledgements

- JavaBayes applet
  - http://www.pmr.poli.usp.br/ltd/Software/javabayes/Home/index.html