Let’s start on BNs…

- Consider $P(X_i)$
  - Assign probability to each $x_i \in \text{Val}(X_i)$
  - Independent parameters $|\text{Val}(X_i)| = k$

- Consider $P(X_1, \ldots, X_n)$
  - How many independent parameters if $|\text{Val}(X_i)| = k$?

\[ k^{n-1} \]
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 \((X \perp Y), \forall X, Y\) subsets of \(\{X_1, \ldots, X_n\}\)

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

- How many independent parameters now?
  - \(n \cdot (k-1)\)

Conditional parameterization – two nodes

- Grade is determined by Intelligence
  - \[ P(I) = \frac{P(I=VH) \cdot \frac{4}{5}}{P(I=VH) \cdot \frac{4}{5} + P(I=AH) \cdot \frac{1}{5}} \]
  - \[ P(G|I) = \frac{P(G=8|I=VH, I=AH) \cdot \frac{1}{5}}{P(G=8|I=VH, I=AH) \cdot \frac{1}{5} + P(G=8|I=VH, I=AH) \cdot \frac{1}{5}} \]
  - \[ P(I=VH, G=8) = P(I=VH) \cdot P(G=8|I=VH) \]
  - \[ P(I=AH) \cdot P(G=8|I=AH) \]
  - \[ = 0.85 \cdot 1 \]
  - \[ = 0.85 \cdot 0.85 \]
  - \[ = 0.085 \cdot 3 \]
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 $(X \perp Y \mid C)$, $\forall X,Y$ 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

- Flu
- Allergy
- Sinus
  - Headache
  - Nose
Number of parameters

Flu -> Sinus
Allergy -> Sinus
Nose -> Sinus
Headache -> Sinus

Key: Independence assumptions

Knowing sinus separates the symptom variables from each other
(Marginal) Independence

- Flu and Allergy are (marginally) independent

<table>
<thead>
<tr>
<th>Flu = t</th>
<th>Flu = f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allergy = t</td>
<td>Allergy = f</td>
</tr>
</tbody>
</table>

- More Generally:

<table>
<thead>
<tr>
<th>Flu = t</th>
<th>Flu = f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allergy = t</td>
<td>Allergy = f</td>
</tr>
</tbody>
</table>

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

$$(X_i \perp \text{NonDescendants}_{X_i} \mid \text{Pa}_{X_i})$$

Explaining away

**Local Markov Assumption:**
A variable $X$ is independent of its non-descendants given its parents and only its parents

$$(X_i \perp \text{NonDescendants}_{X_i} \mid \text{Pa}_{X_i})$$
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:
  \[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \]
- Local Markov Assumption:
  - A variable \( X \) is independent of its non-descendants given its parents and only its parents – (\( X_i \perp \text{NonDescendants}X_i \mid Pa(X_i) \))

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 course 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:**

```
  O --- O
  |    |    
  O   O
```

**Encodes independence assumptions**

If conditional independencies in BN are subset of conditional independencies in \( P \)

**Obtain**

**Joint probability distribution:**

\[
P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i})
\]
Today: The Representation Theorem – BN to Joint Distribution

BN: Encodes independence assumptions

If joint probability distribution:

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \]

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:
  - \( P(C, X_1, \ldots, X_n) = P(C) \prod_i P(X_i \mid C) \)
- 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, $(X \perp Y | C)$, $\forall X,Y$ subsets of $\{X_1, \ldots, X_n\}$

Today: The Representation Theorem

- BN: Encodes independence assumptions
- Joint probability distribution:
  - If conditional independencies in BN are subset of conditional independencies in $P$
    - Obtain $P(X_1, \ldots, X_n) = \prod_{i=1}^n P\left(X_i \mid \text{Pa}_{X_i}\right)$
  - If joint probability distribution:
    - 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 $(X_i \perp \text{NonDescendants}_{X_i} | \text{Pa}_{X_i})$

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

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{Pa}_{X_i})$$
BN Representation Theorem – I-map to factorization

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

$G$ is an I-map of $P$

Joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i})$$

$P$ factorizes according to $G$

**Proof**

$G$ is an I-map of $P$

$P$ factorizes according to $G$

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i})$$

ALL YOU NEED:

Local Markov Assumption:
A variable $X$ is independent of its non-descendants given its parents and only its parents

$$(X_i \perp \text{NonDescendants}_{X_i} | Pa_{X_i})$$
Defining a BN

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

BN Representation Theorem – Factorization to I-map

If joint probability distribution:

\[
P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i})
\]

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:

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{Xi}) \]

Then conditional independencies in BN are subset of conditional independencies in \( P \)

Then conditional independencies in BN are subset of conditional independencies in \( P \)

\( G \) is an I-map of \( P \)

\( P \) factorizes according to \( G \)

Homework 1!!!! 😊

The BN Representation Theorem

If conditional independencies in BN are subset of conditional independencies in \( P \)

important because:
Every \( P \) has at least one BN structure \( G \)

Important because:
Read independencies of \( P \) from BN structure \( G \)

Joint probability distribution:

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{Xi}) \]
Acknowledgements

- JavaBayes applet