CS 188: Artificial Intelligence
Spring 2011

Lecture 15: Bayes’ Nets III
3/16/2011

Pieter Abbeel – UC Berkeley
Many slides over this course adapted from Dan Klein, Stuart Russell, Andrew Moore

Announcements

- Current readings
  - Require login

- Midterm
  - This week’s section will discuss midterm solutions
Our Bayes’ Nets Status

- We now know:
  - What is a Bayes’ net?
  - What joint distribution does a Bayes’ net encode?

- Now: properties of that joint distribution (independence)
  - Key idea: conditional independence
  - Last class: assembled BNs using an intuitive notion of conditional independence as causality
  - Today: formalize these ideas
  - Main goal: answer queries about conditional independence and influence --- important to understand modeling assumptions

- Next: how to compute posteriors quickly (inference)
- Next-next: how to learn a Bayes net from data

Bayes’ Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents’ values
  - CPT: conditional probability table
  - Description of a noisy “causal” process

\[ P(X|a_1 \ldots a_n) \]

\[ P(X|A_1 \ldots A_n) \]

A Bayes net = Topology (graph) + Local Conditional Probabilities
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>~b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>~e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| A  | J  | P(J|A) |
|----|----|------|
| +a | +j | 0.9  |
| +a | ~j | 0.1  |
| ~a | +j | 0.05 |
| ~a | ~j | 0.95 |

| A  | M  | P(M|A) |
|----|----|-------|
| +a | +m | 0.7   |
| +a | ~m | 0.3   |
| ~a | +m | 0.01  |
| ~a | ~m | 0.99  |

| B  | E  | A  | P(A|B,E) |
|----|----|----|----------|
| +b | +e | +a | 0.95     |
| +b | +e | ~a | 0.05     |
| +b | ~e | +a | 0.94     |
| +b | ~e | ~a | 0.06     |
| ~b | +e | +a | 0.29     |
| ~b | +e | ~a | 0.71     |
| ~b | ~e | +a | 0.001    |
| ~b | ~e | ~a | 0.999    |

Size of a Bayes’ Net

- How big is a joint distribution over N Boolean variables?
  \[2^N\]

- How big is an N-node net if nodes have up to k parents?
  \[O(N \times 2^{k+1})\]

- Both give you the power to calculate \(P(X_1, X_2, \ldots, X_n)\)
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (next lecture)
Chain Rule → Bayes net

- Chain rule: can always write any joint distribution as an incremental product of conditional distributions
  \[
P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)
  \]
  \[
P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})
  \]

- Bayes nets: make conditional independence assumptions of the form:
  \[
P(x_i|x_1 \ldots x_{i-1}) = P(x_i|\text{parents}(X_i))
  \]
  giving us:
  \[
P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i))
  \]

Conditional Independence

- Reminder: independence
  - X and Y are independent if
    \[
    \forall x, y \ P(x, y) = P(x)P(y) \quad \rightarrow \quad X \perp Y
    \]
  - X and Y are conditionally independent given Z
    \[
    \forall x, y, z \ P(x, y|z) = P(x|z)P(y|z) \quad \rightarrow \quad X \perp Y|Z
    \]
    equivalently:
    \[
    \forall x, y, z \ P(x|y, z) = P(x|z)
    \]
  - (Conditional) independence is a property of a distribution
Bayes Nets: Assumptions

- Assumptions we are required to make to define the Bayes net when given the graph:
  \[ P(x_i|x_1 \cdots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

- Turns out that probability distributions that satisfy the above ("chain-rule→Bayes net") conditional independence assumptions
  - often can be guaranteed to have many more conditional independences
  - such additional conditional independences can be read off directly from the graph

- Important for modeling: understand assumptions made when choosing a Bayes net graph

Example

- Conditional independence assumptions directly from simplifications in chain rule:

- Additional implied conditional independence assumptions?
Independence in a BN

- Given a Bayes net graph
  - Important question: Are two nodes necessarily independent given certain evidence?
  - If no, can prove with a counter example
    - I.e., pick a set of CPT’s, and show that the independence assumption is violated by the resulting distribution
  - If yes,
    - For now we are able to prove using algebra (tedious in general)
    - Next: “D-separation” (analyzes graph)

D-separation --- Outline

- Study independence properties for triples

- Any complex example can be analyzed using these three canonical cases
Causal Chains

- This configuration is a “causal chain”

\[ P(x, y, z) = P(x)P(y|x)P(z|y) \]

- Is X independent of Z given Y?

\[ P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} = P(z|y) \] Yes!

- Evidence along the chain “blocks” the influence

Common Cause

- Another basic configuration: two effects of the same cause

- Are X and Z independent?

- Are X and Z independent given Y?

\[ P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)} = P(z|y) \] Yes!

- Observing the cause blocks influence between effects.
Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - No: seeing traffic puts the rain and the ballgame in competition as explanation?
  - This is backwards from the other cases
    - Observing an effect activates influence between possible causes.

Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars \( \{Z\} \)?
  - Yes, if X and Y “separated” by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!

- A path is active if each triple is active:
  - Causal chain \( A \rightarrow B \rightarrow C \) where B is unobserved (either direction)
  - Common cause \( A \leftarrow B \rightarrow C \) where B is unobserved
  - Common effect (aka v-structure) \( A \rightarrow B \leftarrow C \) where B or one of its descendants is observed

- All it takes to block a path is a single inactive segment
D-Separation

- Given query \( X_i \perp\perp X_j | \{X_{k_1}, \ldots, X_{k_n}\} \)
- Shade all evidence nodes
- For all (undirected!) paths between and
  - Check whether path is active
    - If active return \( X_i \perp\perp X_j | \{X_{k_1}, \ldots, X_{k_n}\} \)
- (If reaching this point all paths have been checked and shown inactive)
  - Return \( X_i \perp\perp X_j | \{X_{k_1}, \ldots, X_{k_n}\} \)

Example

\( R \perp\perp B \) Yes
\( R \perp\perp B | T \)
\( R \perp\perp B | T' \)
Example

\[ L \perp T' | T \quad \text{Yes} \]
\[ L \perp B \quad \text{Yes} \]
\[ L \perp B | T \]
\[ L \perp B | T' \]
\[ L \perp B | T, R \quad \text{Yes} \]

Example

- **Variables:**
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I’m sad

- **Questions:**
  \[ T \perp D \]
  \[ T \perp D | R \quad \text{Yes} \]
  \[ T \perp D | R, S \]
All Conditional Independences

- Given a Bayes net structure, can run d-separation to build a complete list of conditional independences that are necessarily true of the form

\[ X_i \perp\!\!\!\!\!\!\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \]

- This list determines the set of probability distributions that can be represented.

Possible to have same full list of conditional independence assumptions for different BN graphs?

- Yes!
- Examples:
### Topology Limits Distributions

- Given some graph topology $G$, only certain joint distributions can be encoded.
- The graph structure guarantees certain (conditional) independences.
- (There might be more independence).
- Adding arcs increases the set of distributions, but has several costs.
- Full conditioning can encode any distribution.

![Graph diagrams](image)

### Causality?

- **When Bayes’ nets reflect the true causal patterns:**
  - Often simpler (nodes have fewer parents).
  - Often easier to think about.
  - Often easier to elicit from experts.

- **BNs need not actually be causal**
  - Sometimes no causal net exists over the domain.
  - E.g. consider the variables *Traffic* and *Drips*.
  - End up with arrows that reflect correlation, not causation.

- **What do the arrows really mean?**
  - Topology may happen to encode causal structure.
  - **Topology only guaranteed to encode conditional independence**.
Example: Traffic

- Basic traffic net
- Let’s multiply out the joint

\[
P(R)
\begin{array}{c|c}
 r & 1/4 \\
 \bar{r} & 3/4 \\
\end{array}
\]

\[
P(T|R)
\begin{array}{c|c}
 r & t & 3/4 \\
 \bar{r} & t & 1/4 \\
 \bar{r} & \bar{t} & 1/2 \\
 r & \bar{t} & 1/2 \\
\end{array}
\]

\[
P(T, R)
\begin{array}{c|c|c}
 r & t & 3/16 \\
 r & \bar{t} & 1/16 \\
 \bar{r} & t & 6/16 \\
 \bar{r} & \bar{t} & 6/16 \\
\end{array}
\]

Example: Reverse Traffic

- Reverse causality?

\[
P(T)
\begin{array}{c|c}
 t & 9/16 \\
 \bar{t} & 7/16 \\
\end{array}
\]

\[
P(R|T)
\begin{array}{c|c|c}
 t & r & 1/3 \\
 \bar{r} & 2/3 \\
 \bar{t} & r & 1/7 \\
 \bar{t} & \bar{r} & 6/7 \\
\end{array}
\]

\[
P(T, R)
\begin{array}{c|c|c}
 r & t & 3/16 \\
 r & \bar{t} & 1/16 \\
 \bar{r} & t & 6/16 \\
 \bar{r} & \bar{t} & 6/16 \\
\end{array}
\]
Example: Coins

- Extra arcs don’t prevent representing independence, just allow non-independence

![Graph with nodes X1 and X2 and directed edge X1 -> X2]

- Adding unneeded arcs isn’t wrong, it’s just inefficient

|   | P(X1)   | P(X2)   | P(X1)   | P(X2|X1) |
|---|---------|---------|---------|---------|
| h | 0.5     | h 0.5   | h 0.5   | h | h 0.5 |
| t | 0.5     | t 0.5   | t 0.5   | t | h 0.5 |
|   | h t 0.5 | t t 0.5 | h t 0.5 | t t 0.5 |

Changing Bayes’ Net Structure

- The same joint distribution can be encoded in many different Bayes’ nets
  - Causal structure tends to be the simplest

- Analysis question: given some edges, what other edges do you need to add?
  - One answer: fully connect the graph
  - Better answer: don’t make any false conditional independence assumptions
Example: Alternate Alarm

If we reverse the edges, we make different conditional independence assumptions.

To capture the same joint distribution, we have to add more edges to the graph.

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes’ net’s joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution
Bayes Nets Status

✅ Representation

- Inference
- Learning Bayes Nets from Data