Lecture 14: Graphical Models

Machine Learning
Queens College
Today

• Graphical Models
  – Representing joint probability distributions compactly and graphically by utilizing conditional dependence
Joint Probability Distribution

Representation: What is the joint probability distribution $P(B, E, R, A, N)$ over five binary variables?

• How many states in total?

• How many independent probabilities do we need?

• How many table entries should we sum over to get $P(R=r)$?
Joint Probability Distribution

• Full joint distribution is sufficient to represent the complete domain and to do any type of probabilistic inferences

• Problems: $n$ - number of random variables, $d$ – number of values
  – Space complexity. A full joint distribution requires to remember $O(d^n)$ numbers
  – Inference complexity. Computing queries requires $O(d^n)$ steps
  – Acquisition problem. Who is going to define all the probabilities?
Probability Models

• What if the variables are independent?
• If \( x \) and \( y \) are independent:

\[
p(x, y) = p(x)p(y)
\]

• The original distribution can be factored if \( B, E, A, R, N \) are independent

\[P(B, E, A, R, N) =\]
Conditional Independence

• Independence assumptions are convenient (Naïve Bayes), but rarely true.
• More often some groups of variables are dependent, but others are independent.
• Still others are conditionally independent.
Conditional Independence

• If two variables are conditionally independent.

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

\[ p(x, z) \neq p(x)p(z) \]

• E.g. y = flu?, x = achiness?, z = headache?

\[ x \perp z|y \]
Factorization of a joint

• Assume \( x \perp z \mid y \)

• How do you factorize: \( p(x, y, z) \)

\[ p(x, y, z) = \]
Factorization of a joint

• What if there is no conditional independence?

• How do you factorize: \( p(x, y, z) \)

\[
p(x, y, z) = \]

Bayesian networks

- A Bayesian Network is a graph in which:
  - A set of random variables makes up the nodes in the network.
  - A set of directed links or arrows connects pairs of nodes.
  - Each node has a conditional probability table that quantifies the effects the parents have on the node.
  - Directed, acyclic graph (DAG), i.e. no directed cycles.

A Bayesian network modeling the joint probability distribution $P(B, E, R, A, N)$
Bayesian networks

Conditional probability tables (CPT):

<table>
<thead>
<tr>
<th>Burglary</th>
<th>F</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>T</td>
<td>0.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Earthquake</th>
<th>F</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>T</td>
<td>0.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alarm</th>
<th>F</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>T</td>
<td>0.35</td>
<td>0.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NeighborCall</th>
<th>F</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>T</td>
<td>0.8</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RadioAnnounce</th>
<th>Earthquake</th>
<th>F</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.7</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>0.56</td>
<td>0.44</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
</tr>
<tr>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
</tr>
<tr>
<td>T</td>
</tr>
</tbody>
</table>

| P(A|B,E) |
|--------|
| F      |
| T      |
Independences in BNs

- Three basic independence structures:
Independences in BNs

- **Indirect cause**: Burglary is independent of NeighborCall given Alarm
  - $P(N|A, B) = P(N|A)$
  - $P(N, B|A) = P(N|A)P(B|A)$
Independences in BNs

- **Common cause:** Alarm is independent of RadioAnnounce given Earthquake
  - \( P(A|E, R) = P(A|E) \)
  - \( P(A, R|E) = P(A|E)P(R|E) \)
Independences in BNs

- **Common effect**: Burglary and Earthquake are marginally independent, but become dependant given Alarm
  - $P(B, E) = P(B)P(E)$
  - $P(B, E|A) \neq P(B|A)P(E|A)$
Bayes Ball Algorithm

\[ x_a \perp x_b \mid x_c \]

- Shade nodes \( x_c \)
- Place a “ball” at each node in \( x_a \)
- Bounce balls around the graph according to rules
- If no balls reach \( x_b \), then conditionally independent.
Ten rules of Bayes Ball Theorem
Independences in BNs

- $\text{Earthquake} \perp \text{Burglary} \mid \text{NeighborCall}$?
- $\text{Burglary} \perp \text{NeighborCall}$ ?
- $\text{Burglary} \perp \text{RadioAnnounce} \mid \text{Earthquake}$?
- $\text{Burglary} \perp \text{RadioAnnounce} \mid \text{NeighborCall}$?
Markov Assumption

• Each variable is independent on its non-descendants, given its parents in Bayesian networks

  • $N \perp B, E, R \mid A$
  • $R \perp B, A, N \mid E$
  • ...

![Bayesian network diagram]
Full Joint Distribution in BNs

Product Rule:

\[ P(A, B) = P(A \mid B)P(B) = P(B \mid A)P(A) \]

\[ P(B, E, A, R, N) = \]
Efficient inference

\[ P(R) = \sum_{B,E,A,N} P(B)P(E)P(A|B,E)P(R|E)P(N|A) \]
Undirected Graphs

- What if we allow undirected graphs?
- What do they correspond to?
- Not Cause/Effect, or Trigger/Response, but general dependence

- Example: Image pixels, each pixel is a bernoulli
  - \( P(x_{11}, \ldots, x_{1M}, \ldots, x_{M1}, \ldots, x_{MM}) \)
  - Bright pixels have bright neighbors

- No parents, just probabilities.
- Grid models are called Markov Random Fields
• Undirected separability is easy.
• To check conditional independence of $A$ and $B$ given $C$, check the Graph **reachability** of $A$ and $B$ without going through nodes in $C$
Next Time

• Inference in Graphical Models