CSC242: Intro to AI

Lecture 17
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tue 26 Mar</td>
<td>Bayesian Networks, Exact Inference, Project 3</td>
<td>14.0–14.4.1</td>
</tr>
<tr>
<td>Thu 28 Mar</td>
<td>Quiz; Approximate Inference in BNs</td>
<td>14.5</td>
</tr>
<tr>
<td>Tue 2 Apr</td>
<td>Uncertainty Over Time</td>
<td>15.0–15.3</td>
</tr>
<tr>
<td>Thu 4 Apr</td>
<td>Decision Making</td>
<td>16.0–16.1; 17.0–17.1</td>
</tr>
<tr>
<td>Fri 5 Apr</td>
<td>Project 3 Due</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23:00</td>
<td></td>
</tr>
</tbody>
</table>
Richard Feynman
1918-1988
Bayesian Networks
Probabilistic Inference

- Full joint distribution: intractable as problem grows
- Independence assumptions reduce number of probabilities required to represent full joint distribution (by factoring it)
- Next: Develop a data structure that represents the dependencies among random variables and can be used to compute any full joint distribution
Bayesian Networks
Cavity

Toothache

Catch
Bayesian Networks

Random Variables

- Cavity
- Toothache
- Catch
Bayesian Networks

- Cavity
  - Toothache
  - Catch

“has direct influence on”
Bayesian Networks

- Cavity
  - Toothache
  - Catch

conditionally independent given parents
Bayesian Networks

- Weather
- Toothache
- Cavity
- Catch
Bayesian Networks

\[ P(\text{Toothache} \mid \text{Cavity}) \]

\[ P(\text{Catch} \mid \text{Cavity}) \]

Conditional Probability Distributions
Bayesian Networks

\[
P(\text{Toothache} \mid \text{Cavity})
\]

\[
P(\text{Catch} \mid \text{Cavity})
\]

\[
P(\text{Cavity})
\]

Prior Probability Distribution
Bayesian Networks

• Each node corresponds to a random variable
• There is a link from X to Y if X has a direct influence on Y (no cycles => DAG)
• The node for $X_i$ stores the conditional distribution $P(X_i \mid Parents(X_i))$
• Root nodes store their prior $P(X_i)$
Bayesian Networks
How-To

• Select random variables required to model the domain

• Add links from causes to effects (“directly influences”)

• No cycles (see book)

• Write down conditional probability distributions for $P(X_i \mid Parents(X_i))$
SO WHAT!
THE GOOD, THE MAD, AND THE UGLY

Edited by Stefan Chiara
Probabilistic Inference

• Computing posterior probabilities for propositions given prior probabilities and observed evidence

• or: Given priors and evidence, compute probabilities given evidence (posteriors)
Inference (Single Variable)

\[ P(X \mid e) = \alpha \, P(X, e) = \alpha \sum_y P(X, e, y) \]

- **Query variable** $X$: $\text{Domain}(X) = \{x_1, \ldots, x_m\}$
- **Evidence variables** $E$: $\{E_1, \ldots, E_k\}$
- **Observations** $e$: $\{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$
- **Unobserved variables** $Y$: $\{Y_1, \ldots, Y_l\}$
  
  $\text{Domain}(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

**Full Joint Prob. Dist.**
Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i)) \]
\[ P(\text{toothache}, \text{cavity}, \text{catch}) = P(\text{toothache} | \text{cavity}) P(\text{catch} | \text{cavity}) P(\text{cavity}) \]
\[
P(\neg\text{toothache}, \text{cavity}, \text{catch}) = 
P(\neg\text{toothache}|\text{cavity})P(\text{catch}|\text{cavity})P(\text{cavity})
\]
\[ P(\neg \text{toothache}, \neg \text{cavity}, \text{catch}) = P(\neg \text{toothache} | \neg \text{cavity})P(\text{catch} | \neg \text{cavity})P(\neg \text{cavity}) \]
Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

\[
P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))
\]
Inference in Bayesian Networks

\[ P(X \mid e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y) \]

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i)) \]
Inference in Bayesian Networks

\[
P(X \mid e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y)
\]

\[
= \alpha \sum_{y} \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i))
\]

- A query can be answered from a Bayesian Network by computing sums of products of conditional probabilities from the network.
Alarm
Alarm

**Burglary**

<table>
<thead>
<tr>
<th></th>
<th>$P(B)$</th>
<th>$P(\neg B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.001</td>
<td>1-0.001</td>
</tr>
</tbody>
</table>

**Earthquake**

<table>
<thead>
<tr>
<th></th>
<th>$P(E)$</th>
<th>$P(\neg E)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.002</td>
<td>1-0.002</td>
</tr>
</tbody>
</table>
Alarm

| B | E | P(A|B,E) |
|---|---|---------|
| t | t | 0.95    |
| t | f | 0.94    |
| f | t | 0.29    |
| f | f | 0.001   |

P(B) = 0.001

P(E) = 0.002
| $B$ | $E$ | $P(A|B,E)$ | $P(\neg A|B,E)$ |
|-----|-----|------------|-----------------|
| $t$ | $t$ | 0.95       | 1-.95           |
| $t$ | $f$ | 0.94       | 1-.94           |
| $f$ | $t$ | 0.29       | 1-.29           |
| $f$ | $f$ | 0.001      | 1-.001          |
| Event | $P(B)$ | $P(E)$ | $P(A|B,E)$ |
|-------|--------|--------|------------|
| Burglary | 0.001 | 0.002 |            |
| Earthquake |       |        |            |
| Alarm |        |        |            |
| JohnCalls |       |        |            |
| MaryCalls |       |        |            |

- $P(B) = 0.001$
- $P(E) = 0.002$
- $P(A|B,E)$:
  - $t, t$: 0.95
  - $t, f$: 0.94
  - $f, t$: 0.29
  - $f, f$: 0.001
**Burglary**

\[ P(B) = 0.001 \]

**Earthquake**

\[ P(E) = 0.002 \]

**Alarm**

\[ P(A) \]

- \[ P(A|B) = 0.95 \]  
- \[ P(A|\neg B) = 0.94 \]  
- \[ P(A|E) = 0.29 \]  
- \[ P(A|\neg E) = 0.001 \]

**JohnCalls**

\[ P(J) \]

- \[ P(J|B) = 0.90 \]  
- \[ P(J|\neg B) = 0.05 \]

**MaryCalls**

\[ P(M) \]

- \[ P(M|E) = 0.7 \]  
- \[ P(M|\neg E) = 0.01 \]
P(Burglary | JohnCalls = True, MaryCalls = True) = P(B | j, m)
\[ P(B \mid j, m) = \alpha \sum_e \sum_a P(B, j, m, e, a) \]
\[ P(B \mid j, m) = \alpha \ P(B, j, m) = \alpha \sum_e \sum_a P(B, j, m, e, a) \]

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i)) \]
\[ P(B, E, A, J, M) = \alpha P(B) P(E) P(A | B, E) P(J | A) P(M | A) \]
\[
P(B, E, A, J, M) = \alpha \ P(B) \ P(E) \ P(A | B, E) \ P(J | A) \ P(M | A)
\]
\[ P(B, E, A, j, m) = \alpha P(B) P(E) P(A \mid B, E) P(j \mid A) P(m \mid A) \]

\[
\begin{array}{c|c}
A & P(J \mid A) \\
\hline
\text{t} & 0.90 \\
\text{f} & 0.05 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
B & E & P(A \mid B, E) \\
\hline
\text{t} & \text{t} & 0.95 \\
\text{t} & \text{f} & 0.94 \\
\text{f} & \text{t} & 0.29 \\
\text{f} & \text{f} & 0.001 \\
\end{array}
\]

\[
\begin{array}{c|c}
A & P(M \mid A) \\
\hline
\text{t} & 0.7 \\
\text{f} & 0.01 \\
\end{array}
\]

- **P(B)**: Burglary, \( P(B) = 0.001 \)
- **P(E)**: Earthquake, \( P(E) = 0.002 \)
- **P(j|A)**: John Calls, \( P(j \mid A) = \begin{cases} 0.90 & \text{t} \\ 0.05 & \text{f} \end{cases} \)
- **P(m|A)**: Mary Calls, \( P(m \mid A) = \begin{cases} 0.7 & \text{t} \\ 0.01 & \text{f} \end{cases} \)

**Alarm** node connects to **Burglary** and **Earthquake** nodes.
\[ P(B, E, A, j, m) = \alpha P(B) P(E) P(A | B, E) P(j | A) P(m | A) \]
\[
P(b, E, A, j, m) = \alpha P(b) P(E) P(A | b, E) P(j | A) P(m | A)
\]
\[ P(b, E, A, j, m) = \alpha P(b) P(E) P(A \mid b, E) P(j \mid A) P(m \mid A) \]
\[ P(b, j, m) = \alpha \sum_e \sum_a P(b) P(e) P(a | b, e) P(j | a) P(m | a) \]
\[ P(b, j, m) = \alpha P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a) + \]

<table>
<thead>
<tr>
<th>A</th>
<th>( P(J \mid A) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>0.90</td>
</tr>
<tr>
<td>( f )</td>
<td>0.05</td>
</tr>
</tbody>
</table>

| \( P(B) \) | 0.001 |

| \( P(E) \) | 0.002 |

<table>
<thead>
<tr>
<th>( P(A \mid B, E) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t \mid t )</td>
</tr>
<tr>
<td>( t \mid f )</td>
</tr>
<tr>
<td>( f \mid t )</td>
</tr>
<tr>
<td>( f \mid f )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( P(M \mid A) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
</tr>
<tr>
<td>( f )</td>
</tr>
</tbody>
</table>

- **Burglary** \( P(b) \)
- **Earthquake** \( P(e) \)
- **Alarm** \( P(A \mid b, e) \)
- **JohnCalls** \( P(j \mid a) \)
- **MaryCalls** \( P(m \mid a) \)
\[
P(b, j, m) = \alpha P(b) P(e) P(a | b, e) P(j | a) P(m | a) + \\ P(b) P(e) P(\neg a | b, e) P(j | \neg a) P(m | \neg a) + \\
\]

| \(A\) | \(P(J|A)\) |
|---|---|
| \(t\) | 0.90 |
| \(f\) | 0.05 |

| \(B\) | \(E\) | \(P(A|B,E)\) |
|---|---|---|
| \(t\) | \(t\) | 0.95 |
| \(t\) | \(f\) | 0.94 |
| \(f\) | \(t\) | 0.29 |
| \(f\) | \(f\) | 0.001 |

| \(A\) | \(P(M|A)\) |
|---|---|
| \(t\) | 0.7 |
| \(f\) | 0.01 |
\[ P(b, j, m) = \alpha \, P(b) \, P(e) \, P(a \mid b, e) \, P(j \mid a) \, P(m \mid a) + \]
\[ P(b) \, P(e) \, P(\neg a \mid b, e) \, P(j \mid \neg a) \, P(m \mid \neg a) + \]
\[ P(b) \, P(\neg e) \, P(a \mid b, \neg e) \, P(j \mid a) \, P(m \mid a) + \]
\[ P(b, j, m) = \alpha P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a) + P(b) P(e) P(\neg a \mid b, e) P(j \mid \neg a) P(m \mid \neg a) + P(b) P(\neg e) P(a \mid b, \neg e) P(j \mid a) P(m \mid a) + P(b) P(\neg e) P(\neg a \mid b, \neg e) P(j \mid \neg a) P(m \mid \neg a) \]
\[
P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b)P(e)P(a \mid b, e)P(j \mid a)P(m \mid a)
\]

Factored full joint dist.

Marginalization
\[ P(B \mid j, m) = \alpha \sum \sum P(B, j, m, e, a) \]

\[ P(b \mid j, m) = \alpha \sum \sum P(b)P(e)P(a \mid b, e)P(j \mid a)P(m \mid a) \]

\[ O(n2^n) \]
\[ P(B \mid j, m) = \alpha \prod \prod P(B, j, m) = \alpha \sum \sum P(B, j, m, e, a) \]

\[ P(b \mid j, m) = \alpha \sum \sum P(b)P(e)P(a \mid b, e)P(j \mid a)P(m \mid a) \]

\[ P(b \mid j, m) = \alpha P(b) \sum P(e) \sum P(a \mid b, e)P(j \mid a)P(m \mid a) \]
\[ P(B \mid j, m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle \]
Exact Inference in BNs

Time Complexity: $O(2^n)$

Space Complexity: $O(n)$
Exact Inference in BNs

• Exact inference in BNs is NP-hard

• Can be shown to be as hard as computing the number of satisfying assignments of a propositional logic formula => #P-hard
Bayesian Networks Summary

- Independence assumptions make probabilistic inference easier
- By factoring the joint distribution
- Bayesian Networks encode conditional independence assumptions among random variables
- And store conditional probabilities
Inference in Bayesian Networks

\[ P(X \mid e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y) \]
\[ = \alpha \sum_y \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i)) \]

- Represent your problem as BN: WRITE NO CODE!
Inference in Bayesian Networks

- Exact inference with BNs is still hard
- But we can do approximate inference efficiently (next time)
- We can learn the conditional probabilities required to do inference from data (in a few weeks)
## Syllabus Update

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tue 26 Mar</td>
<td>Bayesian Networks, Exact Inference, Project 3</td>
<td>14.0-14.4.1</td>
</tr>
<tr>
<td>Thu 28 Mar</td>
<td>Quiz; Approximate Inference in BNs</td>
<td>14.5</td>
</tr>
<tr>
<td>Tue 2 Apr</td>
<td>Uncertainty Over Time</td>
<td>15.0-15.3</td>
</tr>
<tr>
<td>Thu 4 Apr</td>
<td>Decision Making</td>
<td>16.0-16.1; 17.0-17.1</td>
</tr>
<tr>
<td>Fri 5 Apr</td>
<td>Project 3 Due</td>
<td></td>
</tr>
</tbody>
</table>
    23:00