Bayesian networks

Chapter 14
Section 1 – 2

Outline

• Syntax
• Semantics
Bayesian networks

• A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

• Syntax:
  – a set of nodes, one per variable
  – a directed, acyclic graph (link ≈ “directly influences”)
  – a conditional distribution for each node given its parents:
    \[ P(X_i | \text{Parents}(X_i)) \]

• In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over \( X_i \) for each combination of parent values

Example

• Topology of network encodes conditional independence assertions:

```
  Weather
     ▼
      ◦ Toothache
         ▼
      ◦ Catch
```

• \textit{Weather} is independent of the other variables
• \textit{Toothache} and \textit{Catch} are conditionally independent given \textit{Cavity}
Example

• I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

• Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

• Network topology reflects "causal" knowledge:
  – A burglar can set the alarm off
  – An earthquake can set the alarm off
  – The alarm can cause Mary to call
  – The alarm can cause John to call

Example contd.

| B | E | P(A|B,E) |
|---|---|--------|
| T | T | .95    |
| T | F | .94    |
| F | T | .29    |
| F | F | .001   |

| A | P(J|A) |
|---|------|
| T | .90  |
| F | .05  |

| A | P(M|A) |
|---|------|
| T | .70  |
| F | .01  |
Compactness

- A CPT for Boolean $X_i$ with $k$ Boolean parents has $2^k$ rows for the combinations of parent values.
- Each row requires one number $p$ for $X_i = \text{true}$ (the number for $X_i = \text{false}$ is just $1-p$).
- If each variable has no more than $k$ parents, the complete network requires $O(n \cdot 2^k)$ numbers.
- I.e., grows linearly with $n$, vs. $O(2^n)$ for the full joint distribution.
- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5-1 = 31$).

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$P(X_1, \ldots, X_n) = \pi_{i=1}^n P(X_i | \text{Parents}(X_i))$$

e.g., $P(j \land m \land a \land \neg b \land \neg e) = P(j | a) P(m | a) P(a | \neg b, \neg e) P(\neg b) P(\neg e) = ?$
Constructing Bayesian networks

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to $n$
   - add $X_i$ to the network
   - select parents from $X_1, \ldots, X_{i-1}$ such that
     $\mathbb{P}(X_i | \text{Parents}(X_i)) = \mathbb{P}(X_i | X_1, \ldots, X_{i-1})$

This choice of parents guarantees:
$\mathbb{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbb{P}(X_i | X_1, \ldots, X_{i-1})$ (chain rule)
$= \prod_{i=1}^{n} \mathbb{P}(X_i | \text{Parents}(X_i))$ (by construction)

Example

- Suppose we choose the ordering $M, J, A, B, E$

$\mathbb{P}(J | M) = \mathbb{P}(J)$?
Example

• Suppose we choose the ordering $M, J, A, B, E$

$P(J \mid M) = P(J)$? No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$?

Example

• Suppose we choose the ordering $M, J, A, B, E$

$P(J \mid M) = P(J)$? No


$P(B \mid A, J, M) = P(B \mid A)$?

$P(B \mid A, J, M) = P(B)$?
Example

• Suppose we choose the ordering M, J, A, B, E

\[ P(J \mid M) = P(J)? \text{ No} \]
\[ P(A \mid J, M) = P(A \mid J)? \text{ No} \]
\[ P(B \mid A, J, M) = P(B \mid A)? \text{ Yes} \]
\[ P(E \mid B, A, J, M) = P(E \mid A)? \text{ No} \]

Example

• Suppose we choose the ordering M, J, A, B, E

\[ P(J \mid M) = P(J)? \text{ No} \]
\[ P(A \mid J, M) = P(A \mid J)? \text{ No} \]
\[ P(B \mid A, J, M) = P(B \mid A)? \text{ Yes} \]
\[ P(E \mid B, A, J, M) = P(E \mid A)? \text{ No} \]
Example contd.

- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct