Bayesian networks

Chapter 14.1–3

Outline

- Syntax
- Semantics
- Parameterized distributions

Syntax:
- A set of nodes, one per variable
- A directed, acyclic graph (link ≈ “directly influences”)
- A conditional probability table (CPT) giving the distribution over $X_i$ for each combination of parent values

Example

- Network topology encodes conditional independence assertions:
  - Weather is independent of the other variables
  - Toothache and Cavity are conditionally independent given Cavity

Example contd.

<table>
<thead>
<tr>
<th>Burglar ($B$)</th>
<th>Earthquake ($E$)</th>
<th>Alarm ($A$)</th>
<th>JohnCalls ($J$)</th>
<th>MaryCalls ($M$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(B)$</td>
<td>$P(E)$</td>
<td>$P(A</td>
<td>B,E)$</td>
<td>$P(J</td>
</tr>
<tr>
<td>.001</td>
<td>.002</td>
<td>.94</td>
<td>.70</td>
<td>.95</td>
</tr>
<tr>
<td>.95</td>
<td>.29</td>
<td>.05</td>
<td>.01</td>
<td>.94</td>
</tr>
<tr>
<td>.001</td>
<td>.94</td>
<td>.05</td>
<td>.01</td>
<td>.95</td>
</tr>
</tbody>
</table>

A simple graphical notation for conditional independence assertions

Example

- If it's cold, 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:
- Burglar ($B$), Earthquake ($E$), Alarm ($A$), JohnCalls ($J$), MaryCalls ($M$)

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
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. $2^5 - 1 = 31$).

Global semantics defines the full joint distribution as the product of the local conditional distributions:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))$$

For example,

$$P(j \land m \land a \land \neg b \land \neg e) = P(j | a) P(m | a) P(a | \neg b, \neg e) P(\neg b) P(\neg e) \approx 0.00063.$$
Example

Suppose we choose the ordering $M, J, A, B, E$. MaryCalls $\Rightarrow$ JohnCalls

$P(J | M) = P(J)$?

No

$P(A | J, M) = P(A | J)$?

No

$P(B | A, J, M) = P(B | A)$?

Yes

$P(E | B, A, J, M) = P(E | A)$?

No

$P(E | B, A, J, M) = P(E | A, B)$?

Yes

Assessing conditional probabilities is hard in noncausal directions. (Causal models and conditional independence seem hardwired for humans!)

Deciding conditional independence is hard in noncausal directions.

Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed

Example contd.

MaryCalls $\Rightarrow$ JohnCalls $\Rightarrow$ Burglary

Yes

No

No

Yes

No

No

No

$P(B | A, J, M) = P(B | A)$?

Yes

$P(E | B, A, J, M) = P(E | A, B)$?

Yes

Example
Example: Car diagnosis

Initial evidence: car won't start

- Established variables (green), "broken, so fix it" variables (orange)
- Hidden variables (gray) ensure sparse structure, reduce parameters

- Lights
- Oil
- Gas
- Starter
- Broken
- Battery age
- Alternator
- Fanbelt
- Broken
- Battery
- Flat
- Gas gauge
- Fuel line
- Blocked
- Oil light
- Battery meter
- Car won't start
- Dipstick

Chapter 14.1–3

Example: Car insurance

- SocioEcon
- Age
- GoodStudent
- ExtraCar
- Mileage
- VehicleYear
- RiskAversion
- SeniorTrain
- DrivingSkill
- MakeModel
- DrivingHist
- DrivQuality
- Antilock
- Airbag
- CarValue
- HomeBase
- AntiTheft
- Theft
- OwnDamage
- PropertyCost
- LiabilityCost
- MedicalCost
- Cushioning
- Ruggedness
- Accident
- OtherCost
- OwnCost

Chapter 14.1–3

Compact conditional distributions

- CPT grows exponentially with number of parents
- CPT becomes infinite with continuous-valued parent or child

Solution: canonical distributions that are defined compactly

Deterministic nodes are the simplest case:

\[ X = f(Parents(X)) \]

for some function \( f \)

E.g., Boolean functions

NorthAmerican ⇔ Canadian ∨ US ∨ Mexican

E.g., numerical relationships among continuous variables

\[
\frac{\partial \text{Level}}{\partial t} = \text{inflow} + \text{precipitation} - \text{outflow} - \text{evaporation}
\]

Chapter 14.1–3

Compact conditional distributions contd.

- Noisy-OR distributions model multiple noninteracting causes

1) Parents \( U_1, \ldots, U_k \) include all causes (can add leak node)
2) Independent failure probability \( q_i \) for each cause alone

\[ P(X | U_1, \ldots, U_j, \neg U_j + 1, \ldots, \neg U_k) = 1 - \prod_{i=1}^j q_i \]

Cold Flu Malaria

\[
\begin{align*}
\text{Fever} & : 0.0 \quad 1.0 \\
\neg \text{Fever} & : 0.9 \quad 0.1
\end{align*}
\]

\[
\begin{align*}
\text{F F F} & : 0.0 \quad 1.0 \\
\neg \text{F F T} & : 0.9 \quad 0.1 \\
\neg \text{T F F} & : 0.0 \quad 1.0 \\
\neg \text{F T T} & : 0.98 \quad 0.02 = 0.02 \\
\neg \text{T F T} & : 0.94 \quad 0.06 = 0.06 \\
\neg \text{T T F} & : 0.88 \quad 0.12 = 0.12 \\
\neg \text{T T T} & : 0.988 \quad 0.012 = 0.012
\end{align*}
\]

Number of parameters linear in number of parents

Continuous variables

- Gaussian density
  \[ P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
- Uniform density
  \[ P(X = x) = \begin{cases} 0.125 & \text{if } 18 \leq x \leq 26 \\ 0 & \text{otherwise} \end{cases} \]

Chapter 14.1–3

Hybrid (discrete+continuous) networks

- Subsidy?
- Buys?
- Harvest
- Cost?

Option 1: discretization—possibly large errors, large CPTs

Option 2: finitely parameterized canonical families

1) Continuous variable, discrete+continuous parents (e.g., Cost)
2) Discrete variable, continuous parents (e.g., Buys?)

Chapter 14.1–3

Continuous variables

- Gaussian density
  \[ P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
- Uniform density
  \[ P(X = x) = \begin{cases} 0.125 & \text{if } 18 \leq x \leq 26 \\ 0 & \text{otherwise} \end{cases} \]

Chapter 14.1–3

Hybrid (discrete+continuous) networks

- Subsidy?
- Buys?
- Harvest
- Cost?

Option 1: discretization—possibly large errors, large CPTs

Option 2: finitely parameterized canonical families

1) Continuous variable, discrete+continuous parents (e.g., Cost)
2) Discrete variable, continuous parents (e.g., Buys?)

Chapter 14.1–3

Continuous variables

- Gaussian density
  \[ P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
- Uniform density
  \[ P(X = x) = \begin{cases} 0.125 & \text{if } 18 \leq x \leq 26 \\ 0 & \text{otherwise} \end{cases} \]
Continuous child variables

Need one conditional density function for child variable given continuous parents, for each possible assignment to discrete parents.

Most common is the linear Gaussian model, e.g.,

$$P(Cost = c | Harvest = h, Subsidy = \text{true}) = \mathcal{N}(a h + b t, \sigma t)(c) = \frac{1}{\sigma t \sqrt{2\pi}} \exp \left(-\frac{1}{2} \frac{(c - (a h + b t))^2}{\sigma^2 t} \right)$$

Mean Cost varies linearly with Harvest, variance is fixed.

Linear variation is unreasonable over the full range but works OK if the likely range of Harvest is narrow.

Chapter 14.1–3 25

All-continuous network with LG distributions ⇒ full joint distribution is a multivariate Gaussian.

Discrete+continuous LG network is a conditional Gaussian network i.e., a multivariate Gaussian over all continuous variables for each combination of discrete variable values.

Chapter 14.1–3 26

Discrete variable w/ continuous parents

Probability of Buys? given Cost should be a "soft" threshold:

$$P(Buys? = \text{false} | Cost = c) = \Phi((-c + \mu) / \sigma)$$

Where Cost is normally distributed over the full range.

Mean Cost varies linearly with Harvest, variance is fixed.

$$\int_{-\infty}^{x} \mathcal{N}(0, 1)(x) \, dx = \Phi(x)$$

Chapter 14.1–3 27

Sigmoid (or logit) distribution also used in neural networks:

$$P(Buys? = \text{true} | Cost = c) = \frac{1}{1 + \exp((-c + \mu) / \sigma)}$$

Sigmoid has similar shape to probit but much longer tails.

Chapter 14.1–3 28

Summary

Bayes nets provide a natural representation for (causally induced) conditional independence. Topology + CPTs = compact representation of joint distribution. Generally easy for (non)experts to construct.

Canonical distributions (e.g., noisy-OR) = compact representation of CPTs. Continuous variables ⇒ parameterized distributions (e.g., linear Gaussian).

Chapter 14.1–3 29

Why the probit?

1. It's sort of the right shape
2. Can view as hard threshold whose location is subject to noise

Continuous variables

Chapter 14.1–3 30