E016350 - Artificial Intelligence

Lecture 10

Reasoning under Uncertainty & Bayesian ML

Bayesian networks

Aleksandra Pizurica

Ghent University
Spring 2025

Overview

- Syntax
- Semantics
- Parameterized distributions

[R&N], Chapter 13

This presentation is based on: S. Russel and P. Norvig: *Artificial Intelligence: A Modern Approach*, (Fourth Ed.), denoted as [R&N] and the resource page http://aima.cs.berkeley.edu/

Bayesian networks

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

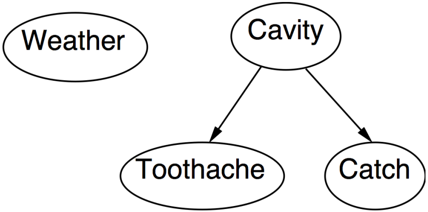
Syntax:

- a set of nodes, one per variable
- a directed, acyclic graph (link $\approx$ "directly influences")
- a conditional distribution for each node given its parents:
  $\mathbf{P}(X_i|parents(X_i))$

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

Network topology encodes conditional independence assertions:



$Weather$ is independent of the other variables
$Toothache$ and $Catch$ are conditionally independent given $Cavity$

Example

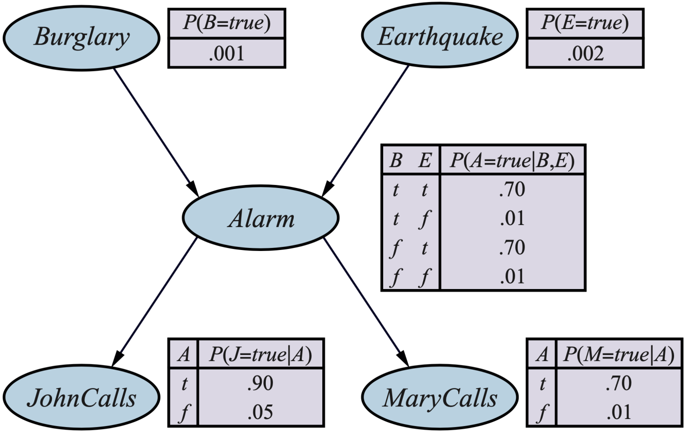
I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$

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.



$P(B=true)$
.001

$P(E=true)$
.002

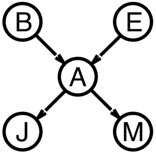
B	E	$P(A=true \mid B,E)$
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	$P(J=true \mid A)$
t	.90
f	.05

A	$P(M=true \mid A)$
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 $X_i = true$
(the number for $X_i = false$ 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. $2^5 - 1 = 31$)

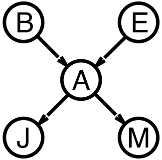
Global semantics

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

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

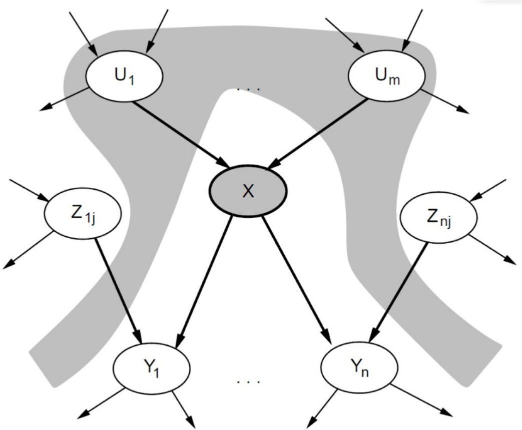
e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$\begin{aligned} &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$



Local semantics

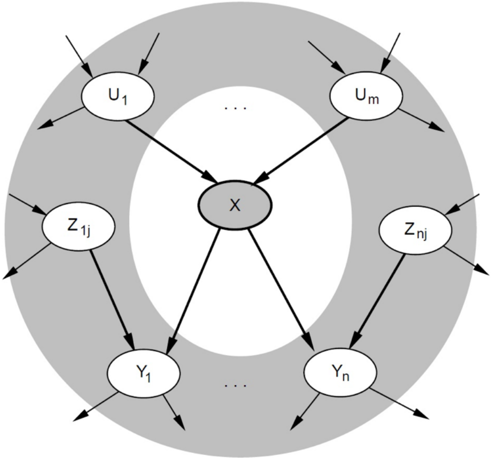
Local semantics: each node is conditionally independent of its nondescendants given its parents



Theorem: Local semantics $\Leftrightarrow$ global semantics

Markov blanket

Each node is conditionally independent of all others given its Markov blanket: parents + children + children's parents:



Constructing Bayesian networks

We need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

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

This choice of parents guarantees the global semantics:

$$\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \quad \text{(by construction)}
\end{aligned}$$

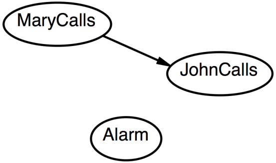
Example

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

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

Example

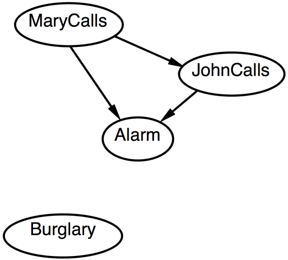
Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$?

Example

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



$P(J|M) = P(J)$? No

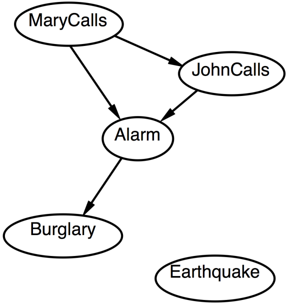
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No

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

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

Example

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



$P(J|M) = P(J)$? No

$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No

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

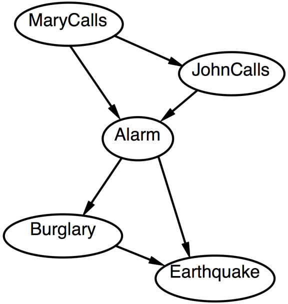
$P(B|A,J,M) = P(B)$? No

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

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

Example

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



$P(J|M) = P(J)$? No

$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No

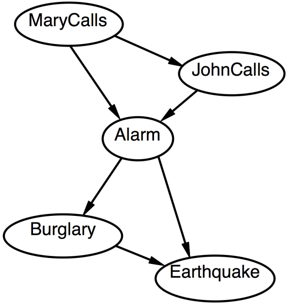
$P(B|A, J, M) = P(B|A)$? Yes

$P(B|A, J, M) = P(B)$? No

$P(E|B, A, J, M) = P(E|A)$? No

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

Example



Compare this network to the one that was given earlier.

Less compact. (Needs $1 + 2 + 4 + 2 + 4 = 13$ numbers)

Deciding conditional independence is hard in noncausal directions.

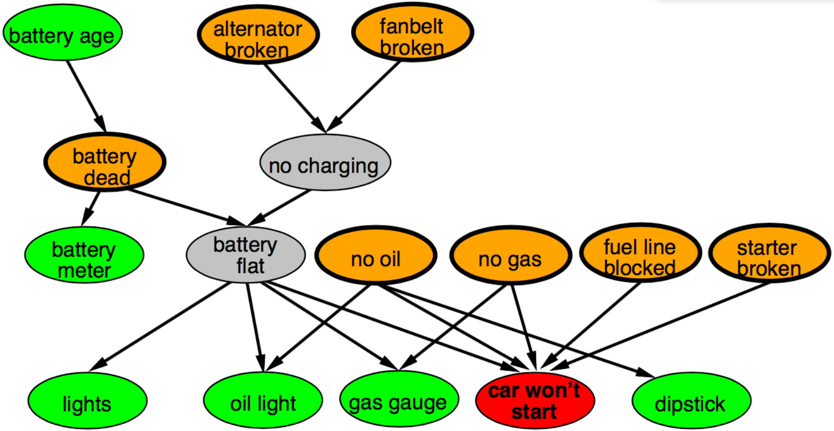
Assessing conditional probabilities is hard in noncausal directions.

Example: Car diagnosis

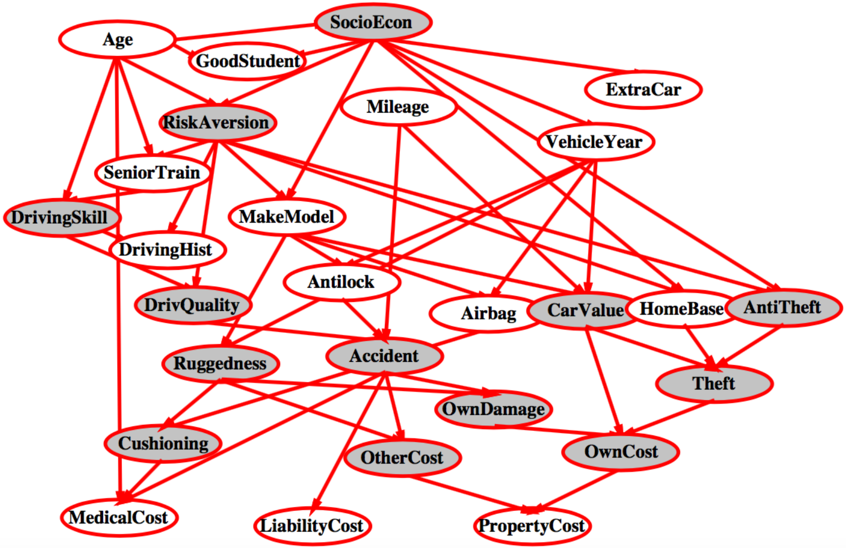
Initial evidence: car won't start

Testable variables (green), "broken, so fix it" variables (orange)

Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car insurance



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
$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$

E.g., numerical relationships among continuous variables

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

Compact conditional distributions contd.

Noisy-OR distributions model multiple noninteracting causes assuming:

1) Parents include all possible causes (can add leak node)
2) Independent failure probability (inhibition probability) q_j for each cause alone

With this, the entire CPT can be built using this general rule:

$$P(x_i|parents(X_i)) = 1 - \prod_{j:X_j=true} q_j$$

Compact conditional distributions contd.

Example: Suppose there are three possible causes for $Fever$,
and these are $Cold$, Flu and $Malaria$. Let their inhibition probabilities be:

$$q_{cold} = P(\neg fever|cold, \neg flu, \neg malaria) = 0.6$$
$$q_{flu} = P(\neg fever|\neg cold, flu, \neg malaria) = 0.2$$
$$q_{malaria} = P(\neg fever|\neg cold, \neg flu, malaria) = 0.1$$

Compact conditional distributions contd.

Example: Suppose there are three possible causes for $Fever$, and these are $Cold$, Flu and $Malaria$. Let their inhibition probabilities be:

$$q_{cold} = P(\neg fever|cold, \neg flu, \neg malaria) = 0.6$$
$$q_{flu} = P(\neg fever|\neg cold, flu, \neg malaria) = 0.2$$
$$q_{malaria} = P(\neg fever|\neg cold, \neg flu, malaria) = 0.1$$

Noisy-OR model: $P(x_i|parents(X_i)) = 1 - \prod_{j:X_j=true} q_j$ yields the CPT:

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	$\mathbf{0.0}$	1.0
F	F	T	0.9	$\mathbf{0.1}$
F	T	F	0.8	$\mathbf{0.2}$
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	$\mathbf{0.6}$
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Compact conditional distributions contd.

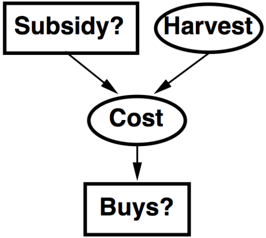
$$P(x_i|parents(X_i)) = 1 - \prod_{j:X_j=true} q_j$$

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	$\mathbf{0.0}$	1.0
F	F	T	0.9	$\mathbf{0.1}$
F	T	F	0.8	$\mathbf{0.2}$
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	$\mathbf{0.6}$
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

The number of parameters increases linearly with the number of parents.
($O(k)$ parameters instead of $O(2^k)$ for the full CPT)

Hybrid (discrete+continuous) networks

Discrete ($Subsidy?$ and $Buys?$); continuous ($Harvest$ and $Cost$)



Option 1: discretization—possibly large errors, large CPTs
Option 2: finitely parameterized canonical families

Two new types of distributions to specify:

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

Continuous child variables

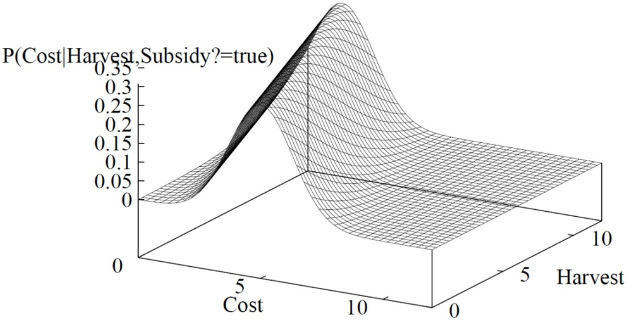
How to construct conditional density function for child variable given continuous parents, for each possible assignment to discrete parents?

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

$$
\begin{aligned}
&P(Cost = c | Harvest = h, Subsidy? = true) \\
&= \quad N(a_t h + b_t, \sigma_t)(c) \\
&= \quad \frac{1}{\sigma_t \sqrt{2\pi}} exp \left(-\frac{1}{2} \left(\frac{c - (a_t h + b_t)}{\sigma_t} \right)^2 \right)
\end{aligned}
$$

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

Continuous child variables

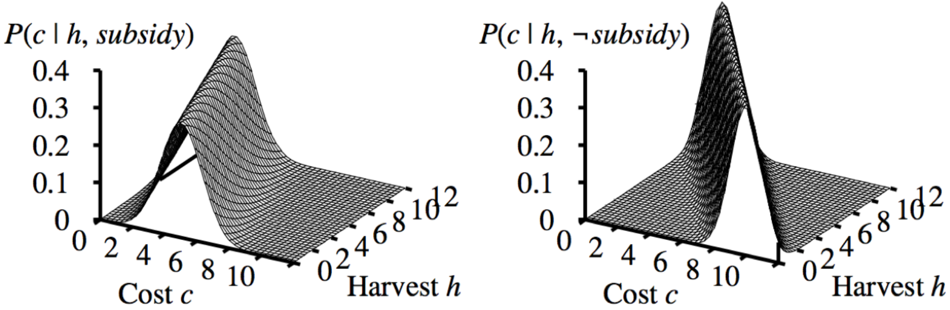


All-continuous network with LG distributions

$\implies$ full joint distribution is a multivariate Gaussian

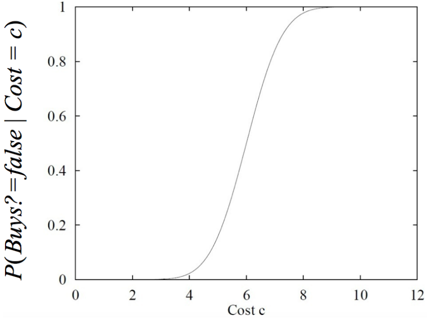
Discrete+continuous LG network is a conditional Gaussian network i.e., a multivariate Gaussian over all continuous variables for each combination of discrete variable values

Continuous child variables



Discrete variable with continuous parents

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

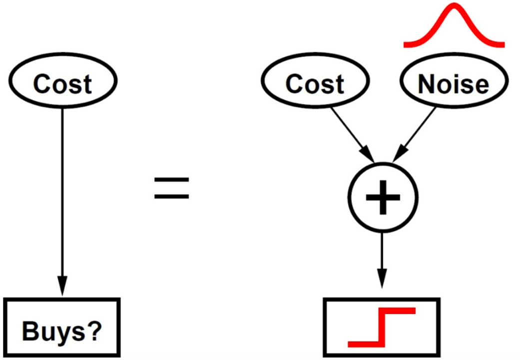


Probit distribution uses integral of Gaussian:

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

$P(Buys?\!=\!true|Cost\!=\!c) = \Phi((-c+\mu)/\sigma)$

Why the probit?



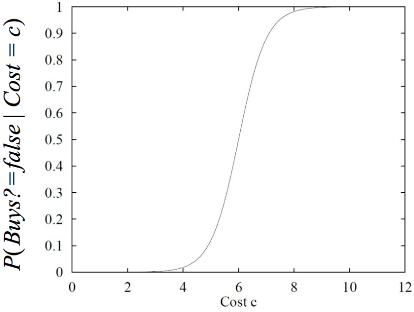
1. It's sort of the right shape
2. Can be viewed as hard threshold whose location is subject to noise

Discrete variable contd.

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

$$P(Buys? = true|Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$

Sigmoid has similar shape to probit but much longer tails:



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 $\implies$ parameterized distributions (e.g., linear Gaussian)