Lecture 11
Bayesian Networks

M.Filippone

Artificial Intelligence 4 (AI4)
School of Computing Science
University of Glasgow (UK)

Lecture 11: Reference

This lecture corresponds to Chapter 14 of the following textbook:

Artificial Intelligence. A Modern Approach

Stuart Russell and Peter Norvig

aima.cs.berkeley.edu/

Inference from joint distributions

- Joint distributions

$$\mathbf{P}(X_1, \ldots, X_n)$$

- Allow to answer all inferential question
  - Marginal distributions
  - Conditional (posterior) distributions

Inference from joint distributions

Let d be the number of states - problems:

1. Time complexity $O(d^n)$
2. Space complexity $O(d^n)$
3. How to find the numbers for $O(d^n)$ entries?

Chain Rule

- Let's analyze the joint distribution using the chain rule:

$$\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_n | X_1, \ldots, X_{n-1}) \mathbf{P}(X_1, \ldots, X_{n-1}) \\
&= \mathbf{P}(X_n | X_1, \ldots, X_{n-1}) \times \\
&\quad\; \mathbf{P}(X_{n-1} | X_1, \ldots, X_{n-2}) \mathbf{P}(X_1, \ldots, X_{n-2}) \\
&= \ldots \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1})
\end{aligned}$$

- Can we exploit this?

Outline

Conditional independence

Belief Networks

Constructing belief networks

Outline

Conditional independence

Belief Networks

Constructing belief networks

Independence

Two random variables A and B are independent iff

- $P(A|B) = P(A)$
- or $P(A, B) = P(A|B)P(B) = P(A)P(B)$

For example: A and B are two coin tosses

Independence

- If n variables are independent, the full joint is

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_i \mathbf{P}(X_i)$$

- Hence can be specified by just n numbers
- Not a very useful assumption

Conditional independence

Consider the following problem with three random variables:

- *Toothache* (T)
- *Cavity* (C)
- *Probe catch* (P) (probe catches in my tooth)

The full joint distribution has $2^3 - 1 = 7$ independent entries

Conditional independence

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

$$P(P|T, C) = P(P|C)$$

- In other words, P is conditionally independent of T given C

Conditional independence

- Other equivalent statements:

$$P(T|P,C) = P(T|C)$$

$$P(T,P|C) = P(T|C)P(P|C)$$

- Why?

Conditional independence

- Why is $P(T|P,C) = P(T|C)$?

$$
\begin{aligned}
P(T|P,C) &= P(P|T,C)P(T|C)/P(P|C) \\
&= P(P|C)P(T|C)/P(P|C) \\
&= P(T|C)
\end{aligned}
$$

- We used

$$P(P|T,C) = P(P|C)$$

Conditional independence

- Why is $P(T, P|C) = P(T|C)P(P|C)$?

$$
\begin{aligned}
P(T, P|C) &= P(T|P, C)P(P|C) \\
&= P(T|C)P(P|C)
\end{aligned}
$$

- We used the product rule and

$$P(T|P, C) = P(T|C)$$

Conditional independence

- The same independence holds if I haven't got a cavity:

$$P(P|T, \neg C) = P(P|\neg C)$$

Conditional independence

- Full joint distribution can now be written as

$$\mathbf{P}(T, P, C) = \mathbf{P}(T, P|C)\mathbf{P}(C) = \mathbf{P}(T|C)\mathbf{P}(P|C)\mathbf{P}(C)$$

- I.e., 2 + 2 + 1 = 5 independent numbers

Outline

Conditional independence

Belief Networks

Constructing belief networks

Belief networks

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

Belief networks

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 Boolean case, conditional distribution represented as a conditional probability table (CPT)

Example

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

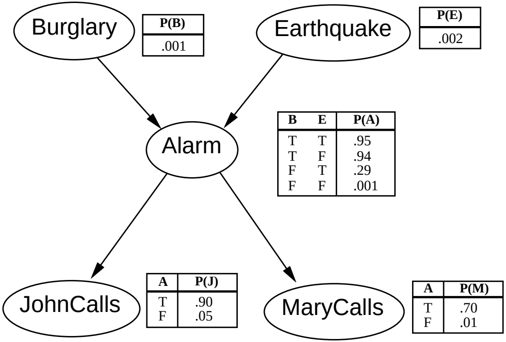
Example

Variables:

- *Burglar* (B)
- *Earthquake* (E)
- *Alarm* (A)
- *John Calls* (J)
- *Mary Calls* (M)

Example

Network topology reflects causal knowledge:



Note: $\leq k$ parents $\Rightarrow O(d^k n)$ numbers vs $O(d^n)$

Semantics

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

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

- For example

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$

is given by?

Semantics

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

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

- For example

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$

is given by?

$$P(\neg B)P(\neg E)P(A|\neg B \wedge \neg E)P(J|A)P(M|A)$$

Semantics

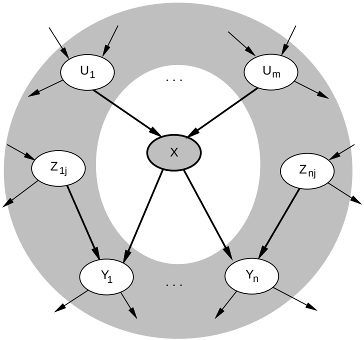
- "Local" semantics: each node is conditionally independent of its non-descendants given its parents

Local semantics $\Leftrightarrow$ Global semantics

Markov blanket

- Each node is conditionally independent of all others given its Markov blanket.

Markov blanket $=$ parents $+$ children $+$ children′s parents



Constructing belief networks

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

Outline

Conditional independence

Belief Networks

Constructing belief networks

Constructing belief networks

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   2.1 Add X_i to the network
   2.2 Select parents from $X_1, \ldots, X_{i-1}$ such that

$$\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1,\ \ldots,\ X_{i-1})$$

Constructing belief networks

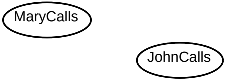
- This choice of parents guarantees the global semantics
- Applying the chain rule:

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) && (1) \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) && (2)
\end{aligned}
$$

by construction

Example - Choosing wrong ordering

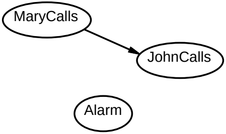
Suppose we choose the ordering *M*, *J*, *A*, *B*, *E*



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

Example - Choosing wrong ordering

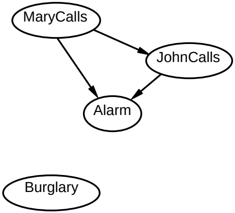
Suppose we choose the ordering M, J, A, B, E



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

Example - Choosing wrong ordering

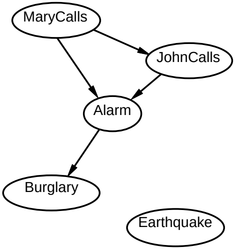
Suppose we choose the ordering M, J, A, B, E



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

Example - Choosing wrong ordering

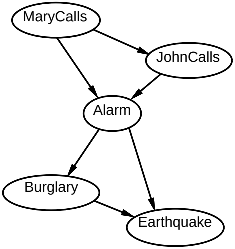
Suppose we choose the ordering M, J, A, B, E



- $P(B|A, J, M) = P(B|A)$? 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 - Choosing wrong ordering

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

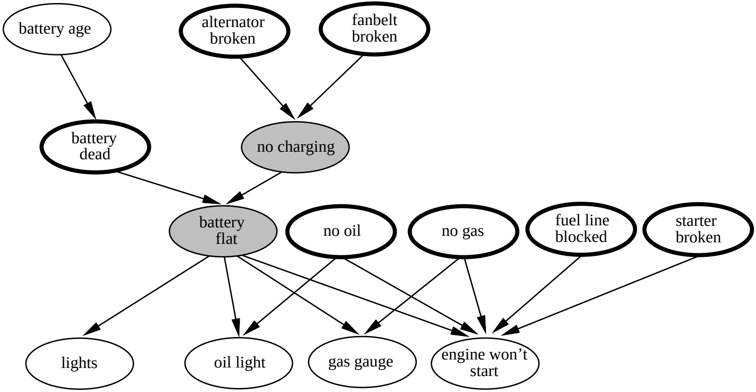


- $P(E|B, A, J, M) = P(E|A)$? No
- $P(E|B, A, J, M) = P(E|A, B)$? Yes

Example: Car diagnosis

- Initial evidence: engine won't start
- Testable variables (thin ovals)
- Diagnosis variables (thick ovals)
- Hidden variables (shaded) ensure sparse structure, reduce parameters

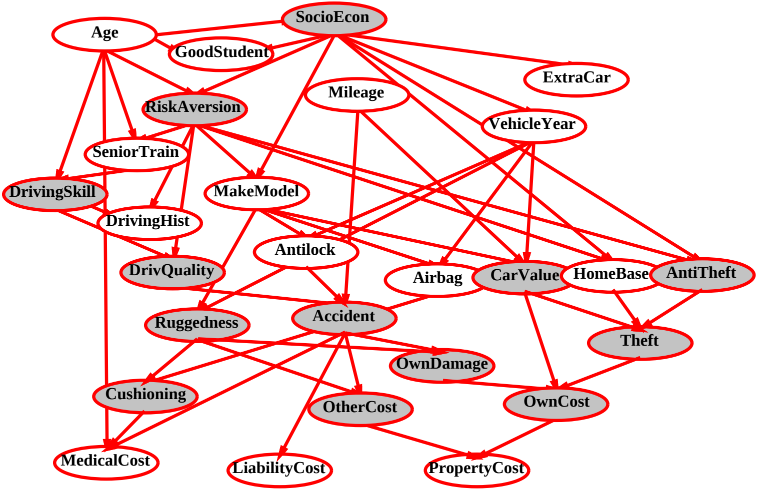
Example: Car diagnosis



Example: Car insurance

- Predict claim costs (medical, liability, property) given data on application form (other unshaded nodes)

Example: Car insurance



End of the Lecture