Belief networks

Chapter 15

Outline

- Conditional independence
- Bayesian networks: syntax and semantics
- Exact inference
- Approximate inference
Independence

Two random variables \( A, B \) are (absolutely) independent iff

\[
P(A|B) = P(A)
\]
or

\[
P(A, B) = P(A|B)P(B) = P(A)P(B)
\]
e.g., \( A \) and \( B \) are two coin tosses

If \( n \) Boolean variables are independent, the full joint is

\[
P(X_1, \ldots, X_n) = \Pi_i P(X_i)
\]
hence can be specified by just \( n \) numbers

Absolute independence is a very strong requirement, seldom met

---

Conditional independence

Consider the dentist problem with three random variables:

- Toothache, Cavity, Catch (steel probe catches in my tooth)

The full joint distribution has \( 2^3 - 1 = 7 \) independent entries

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

1. \( P(\text{Catch}|\text{Toothache}, \text{Cavity}) = P(\text{Catch}|\text{Cavity}) \)
i.e., Catch is conditionally independent of Toothache given Cavity

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

2. \( P(\text{Catch}|\text{Toothache}, \neg\text{Cavity}) = P(\text{Catch}|\neg\text{Cavity}) \)
Conditional independence contd.

Equivalent statements to (1)

(1a) \( P(\text{Toothache}|\text{Catch, Cavity}) = P(\text{Toothache}|\text{Cavity}) \) Why??

(1b) \( P(\text{Toothache, Catch}|\text{Cavity}) = P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity}) \) Why??

Full joint distribution can now be written as

\[
P(\text{Toothache, Catch, Cavity}) = P(\text{Toothache, Catch}|\text{Cavity})P(\text{Cavity})
= P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity})
\]

i.e., \( 2 + 2 + 1 = 5 \) independent numbers (equations 1 and 2 remove 2)

---

Conditional independence contd.

Equivalent statements to (1)

(1a) \( P(\text{Toothache}|\text{Catch, Cavity}) = P(\text{Toothache}|\text{Cavity}) \) Why??

\[
P(\text{Toothache}|\text{Catch, Cavity})
= P(\text{Catch}|\text{Toothache, Cavity})P(\text{Toothache}|\text{Cavity})P(\text{Cavity})P(\text{Catch}|\text{Cavity})

= P(\text{Catch}|\text{Toothache, Cavity})P(\text{Toothache}|\text{Cavity})P(\text{Cavity}) \text{ (from 1)}

= P(\text{Toothache}|\text{Cavity})
\]

(1b) \( P(\text{Toothache, Catch}|\text{Cavity}) = P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity}) \) Why??

\[
P(\text{Toothache, Catch}|\text{Cavity})
= P(\text{Toothache}|\text{Catch, Cavity})P(\text{Catch}|\text{Cavity}) \text{ (product rule)}

= P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity}) \text{ (from 1a)}
\]
Belief 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:
  \[ P(X_i | \text{Parents}(X_i)) \]

In the simplest 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 it's set off by minor earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects "causal" knowledge:

Note: $\leq k$ parents $\Rightarrow O(d^k n)$ numbers vs. $O(d^n)$
"Global" semantics defines the full joint distribution as the product of the local conditional distributions:

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{Parents}(X_i)) \]

e.g., \( P(J \land M \land A \land \neg B \land \neg E) \) is given by:

\[ \emptyset \]

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

Theorem: Local semantics \( \iff \) global semantics
**Markov blanket**

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

![Diagram of Markov blanket]

**Constructing belief networks**

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

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

This choice of parents guarantees the global semantics:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^n P(X_i|X_1, \ldots, X_{i-1}) \quad \text{(chain rule)}$$

$$= \prod_{i=1}^n P(X_i|\text{Parents}(X_i)) \quad \text{by construction}$$
Example

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

- $P(J|M) = P(J)\,?\,\,\,\text{NO}$
- $P(A|J,M) = P(A|J)\,?\,\,\,\text{NO}$
- $P(A|J,M) = P(A)\,?\,\,\,\text{YES}$
- $P(B|A, J, M) = P(B)\,?\,\,\,\text{NO}$
- $P(E|B, A, J, M) = P(E|A)\,?\,\,\,\text{NO}$
- $P(E|B, A, J, M) = P(E|A, B)\,?\,\,\,\text{YES}$

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 insurance

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

Compact conditional distributions

CPT grows exponentially with no. of parents
CPT becomes infinite with continuous-valued parent or child

Solution: canonical distributions that are defined compactly

Deterministic nodes are the simplest case:
\[ X = f(\text{Parents}(X)) \] for some function \( f \)

E.g., Boolean functions
\[ \text{NorthAmerican} \leftrightarrow \text{Canadian} \lor \text{US} \lor \text{Mexican} \]

E.g., numerical relationships among continuous variables
\[ \frac{\partial \text{Level}}{\partial t} = \text{inflow} + \text{precipitation} - \text{outflow} - \text{evaporation} \]
Inference in Belief Networks

- Compute posterior probability distribution for a set of query variables given exact values for some evidence variables – $P(\text{Query}|\text{Evidence})$
- Ex: $P(\text{Burglary}|\text{JohnCalls})$
- Belief networks allow any node to be query / evidence
- Belief-Net-Ask – function for computing probability for given query variable

---

**Inference tasks**

Simple queries: compute posterior marginal $P(X|E=e)$
  
  e.g., $P(\text{NoGas}|\text{Gauge = empty}, \text{Lights = on}, \text{Starts = false})$

Conjunctive queries: $P(X_1, X_2|E=e) = P(X_1|E=e)P(X_2|X_1, E=e)$

Optimal decisions: decision networks include utility information;
probabilistic inference required for $P(\text{outcome}|\text{action, evidence})$

**Value of information:** which evidence to seek next?

**Sensitivity analysis:** which probability values are most critical?

**Explanation:** why do I need a new starter motor?
### Section 15.4 – Inference in Multiply Connected Belief Networks

**Function**: `BeliefNetAsk(X)` returns a probability distribution over the values of X.

**Input**: X, a random variable.

**Support**

**Function**: `SupportExcept(X, E)` returns $P(X | E_\neg X)$.

- If `Evidence(X)` then return observed point distribution for X.
- Else:
  - Calculate $P(E_\neg X | X)$ = `EvidenceExcept(X, E)`.
  - Let $U = \text{Parents}(X)$.
  - If $U$ is empty, then return $P(X | E_\neg X)$.
  - Else:
    - For each $U_i \in U$:
      - Calculate and store $P(U_i | E_\neg X, X)$ = `SupportExcept(U_i, X)`.
  - Return $P(X | E_\neg X) \prod_{i} P(U_i | E_\neg X, X)$.

**Function**: `EvidenceExcept(X, Y)` returns $P(E_\neg X | Y)$.

- Y = Children[X].
- If Y is empty, then return a uniform distribution.
- Else:
  - For each Y_i in Y:
    - Calculate $P(E_\neg X | Y)$ = `EvidenceExcept(Y_i, null)`.
    - Let $Z_i = \text{Parents}(Y_i)$.
    - For each $Z_j$ in $Z_i$:
      - Calculate $P(Z_j | E_\neg X, Y_i)$ = `SupportExcept(Z_j, Y_i)`.
  - Return $\prod_{i} P(E_\neg X | Y) \prod_{j} P(Z_j | E_\neg X, Y_i)$.

Backward chaining algorithm for solving probabilistic queries on a polytree (at most one undirected path between any two nodes in the network).

$P(X | E_\neg X, Y)$

- Probability of X given evidence E, where E is separated by conditional independence.

$P(E_\neg X | Y)$

- Probability of E given evidence X, where E is separated by conditional independence.
Section 15.6 – Other Approaches to Uncertain Reasoning