Reasoning Under Uncertainty: Belief Network Inference

CPSC 322 Lecture 27

March 21, 2007
Textbook §9.4

Lecture Overview

1. Recap
2. Observing Variables
3. Belief Network Inference
4. Factors

Components of a belief network

A belief network consists of:

- a directed acyclic graph with nodes labeled with random variables
- a domain for each random variable
- a set of conditional probability tables for each variable given its parents (including prior probabilities for nodes with no parents).

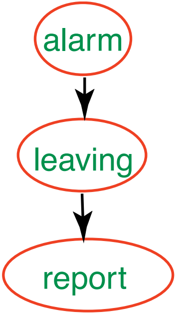
How to construct a belief network

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule): $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})$
- The parents pX_i of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is, $pX_i \subseteq X_1, \ldots, X_{i-1}$ and $P(X_i | pX_i) = P(X_i | X_1, \ldots, X_{i-1})$
- So $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | pX_i)$

Lecture Overview

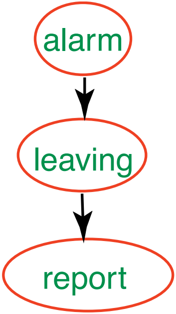
1. Recap
2. Observing Variables
3. Belief Network Inference
4. Factors

Chain



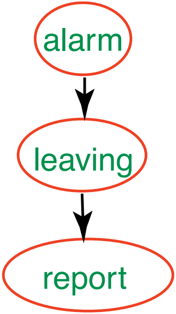
- $alarm$ and $report$ are independent:

Chain



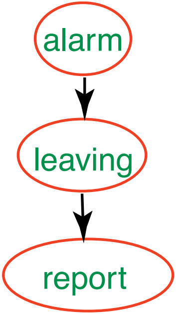
- $alarm$ and $report$ are independent: false.

Chain



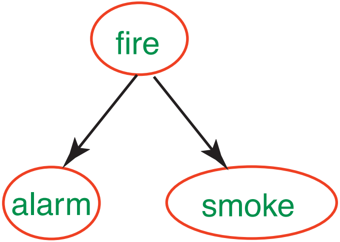
- $alarm$ and $report$ are independent: false.
- $alarm$ and $report$ are independent given $leaving$:

Chain



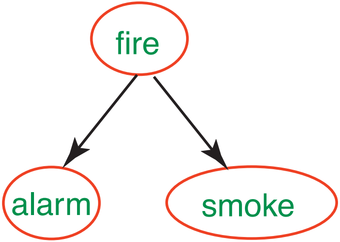
- *alarm* and *report* are independent: false.
- *alarm* and *report* are independent given *leaving*: true.
- Intuitively, the only way that the *alarm* affects *report* is by affecting *leaving*.

Common ancestors



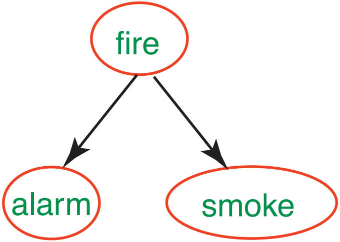
- *alarm* and *smoke* are independent:

Common ancestors



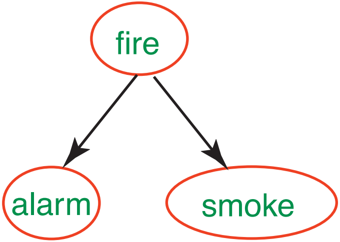
- $alarm$ and $smoke$ are independent: false.

Common ancestors



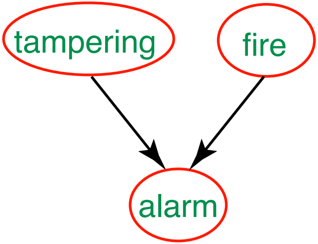
- $alarm$ and $smoke$ are independent: false.
- $alarm$ and $smoke$ are independent given $fire$:

Common ancestors



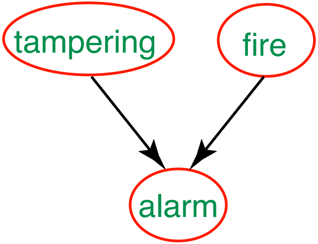
- $alarm$ and $smoke$ are independent: false.
- $alarm$ and $smoke$ are independent given $fire$: true.
- Intuitively, $fire$ can explain $alarm$ and $smoke$; learning one can affect the other by changing your belief in $fire$.

Common descendants



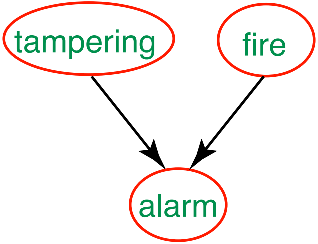
- $tampering$ and $fire$ are independent:

Common descendants



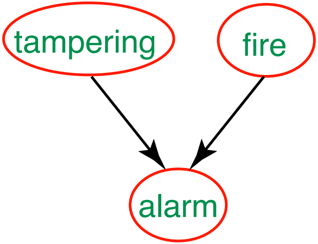
- $tampering$ and $fire$ are independent: true.

Common descendants



- $tampering$ and $fire$ are independent: true.
- $tampering$ and $fire$ are independent given $alarm$:

Common descendants



- $tampering$ and $fire$ are independent: true.
- $tampering$ and $fire$ are independent given $alarm$: false.
- Intuitively, $tampering$ can explain away $fire$

Lecture Overview

1. Recap
2. Observing Variables
3. Belief Network Inference
4. Factors

Belief Network Inference

- Our goal: compute probabilities of variables in a belief network
- Two cases:
  1. the unconditional (prior) distribution over one or more variables
  2. the posterior distribution over one or more variables, conditioned on one or more observed variables

Evidence

- If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \ldots \wedge Y_j = v_j$:

$$
\begin{aligned}
& P(Z|Y_1 = v_1, \ldots, Y_j = v_j) \\
= \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{P(Y_1 = v_1, \ldots, Y_j = v_j)} \\
= \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{\sum_Z P(Z, Y_1 = v_1, \ldots, Y_j = v_j).}
\end{aligned}
$$

- So the computation reduces to the probability of $P(Z, Y_1 = v_1, \ldots, Y_j = v_j)$.

Belief Network Inference

- Our goal: compute probabilities of variables in a belief network
- Two cases:
  1. the unconditional (prior) distribution over one or more variables
  2. the posterior distribution over one or more variables, conditioned on one or more observed variables
- To address both cases, we only need a computational solution to case 1
- Our method: exploiting the structure of the network to efficiently eliminate (sum out) the non-observed, non-query variables one at a time.

Lecture Overview

1. Recap
2. Observing Variables
3. Belief Network Inference
4. Factors

Factors

- A factor is a representation of a function from a tuple of random variables into a number.
- We will write factor f on variables $X_1, \ldots, X_j$ as $f(X_1, \ldots, X_j)$.
- A factor denotes a distribution over the given tuple of variables in some (unspecified) context
  - e.g., $P(X_1, X_2)$ is a factor $f(X_1, X_2)$
  - e.g., $P(X_1, X_2, X_3 = v_3)$ is a factor $f(X_1, X_2)$
  - e.g., $P(X_1, X_3 = v_3 | X_2)$ is a factor $f(X_1, X_2)$

Manipulating Factors

- We can make new factors out of an existing factor
- Our first operation: we can assign some or all of the variables of a factor.
  - $f(X_1 = v_1, X_2, \ldots, X_j)$, where $v_1 \in dom(X_1)$, is a factor on $X_2, \ldots, X_j$.
  - $f(X_1 = v_1, X_2 = v_2, \ldots, X_j = v_j)$ is a number that is the value of f when each X_i has value v_i.
- The former is also written as $f(X_1, X_2, \ldots, X_j)_{X_1 = v_1, \ldots, X_j = v_j}$

Example factors

$r(X, Y, Z)$:

X	Y	Z	val
t	t	t	0.1
t	t	f	0.9
t	f	t	0.2
t	f	f	0.8
f	t	t	0.4
f	t	f	0.6
f	f	t	0.3
f	f	f	0.7

$r(X{=}t, Y, Z)$:

Y	Z	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$r(X{=}t, Y, Z{=}f)$:

Y	val
t	0.9
f	0.8

$r(X{=}t, Y{=}f, Z{=}f) = 0.8$