Bayes III
(Introduction to Bayes Nets)

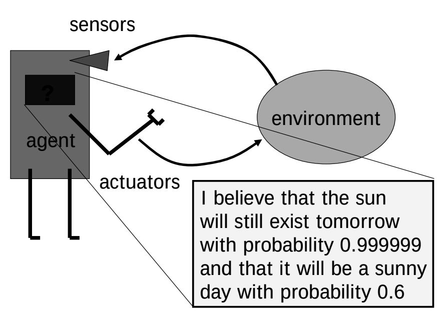
Lecture #18

11/04/08

Announcements

- Programming Assignment #2 is now due
  - Any comments?
- ACM meeting tomorrow: 5pm USC100
  - how to get into the world of computer graphics
- Election day : vote!
  - Unless you voted already (this isn't Chicago…)

Probabilistic Agent



Problem

- At a certain time t, the KB of an agent is some collection of beliefs
  - Every belief has a degree of certainty
- At time *t* the agent's sensors make an observation that changes the strength of one of its beliefs
- How should the agent update the strength of its other beliefs?

Purpose of Bayesian Networks

- Facilitate the description of a collection of beliefs by
  - making causality relations explicit
  - exploiting conditional independence
- Provide efficient methods for:
  - Representing a joint probability distribution
  - Updating belief strengths when new evidence is observed

Review

- Conditional Probability: P(A|B)
  - The probability of A given that B is true
  - The frequency with which A will be observed, considering only samples in which B is true
  - E.g. P(toothache | cavity) *from last lecture*
- Bayes Rule

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)}$$

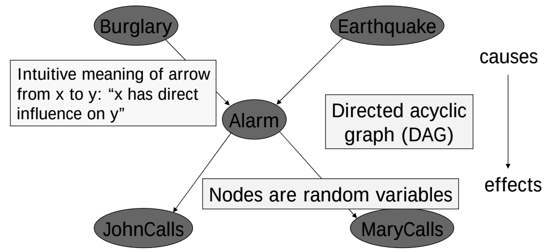
Example

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

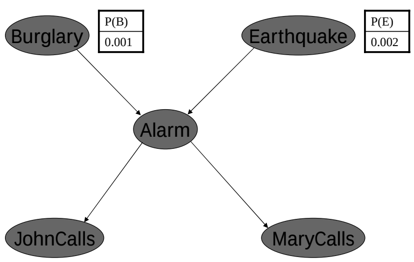
Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects "causal" knowledge:
- A burglary 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

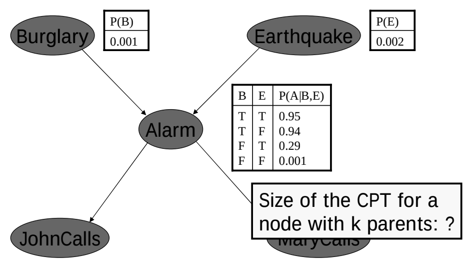
A Simple Network



Assigning Probabilities to Roots

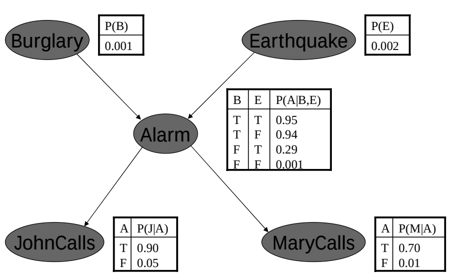


Conditional Probability Tables



B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

Conditional Probability Tables

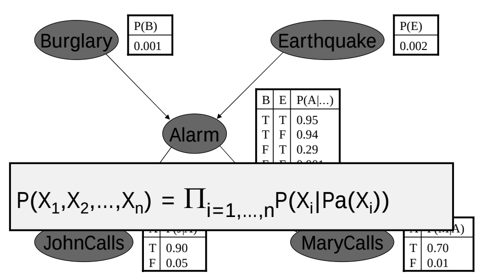


B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

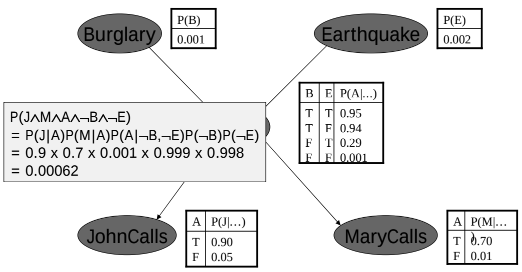
A	P(M|A)
T	0.70
F	0.01

What the BN Means



$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1,\ldots,n} P(X_i \mid Pa(X_i))$$

Calculation of Joint Probability



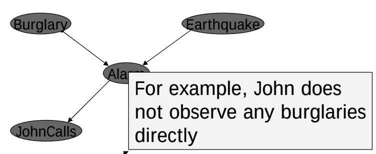
Independence of Random Variables

- Two variables X and Y are **independent** if
  - $P(X = x|Y = y) = P(X = x)$ for all values x,y
  - That is, learning the values of Y does not change knowledge of X
- If X and Y are independent then
  - $P(X,Y) = P(X|Y)P(Y) = P(X)P(Y)$
- In general, if $X_1,\ldots,X_n$ are independent, then
  - $P(X_1,\ldots,X_n) = P(X_1)\ldots P(X_n)$
  - Requires O(n) parameters

Conditional independence

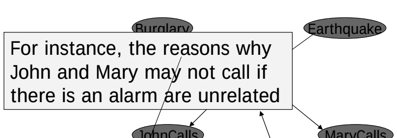
- Unfortunately, most random variables of interest are not independent
- A more suitable notion is that of conditional independence
- Two variables X and Y are **conditionally independent** given Z if
  - $P(X = x|Y = y,Z=z) = P(X = x|Z=z)$ for all values x,y,z
  - That is, learning the values of Y does not change prediction of X once we know the value of Z
  - Notation: $I(X ; Y | Z)$

What the BN Encodes



- Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or ¬Alarm
- The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬Alarm

What the BN Encodes

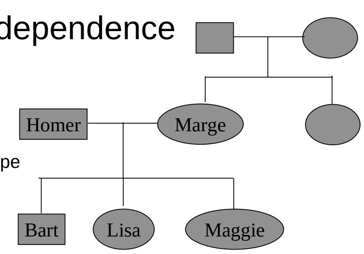


- Each of the beliefs ... Alarm or ¬Alarm
- The beliefs JohnCalls nd MaryCalls are ndependent given Alarm r ¬Alarm

Note that these reasons could be other beliefs in the network. The probabilities summarize these non-explicit beliefs

Example of conditional independence

- A node represents an individual's genotype
- Ancestors affect descendants' genotype by passing genetic material through intermediate generations

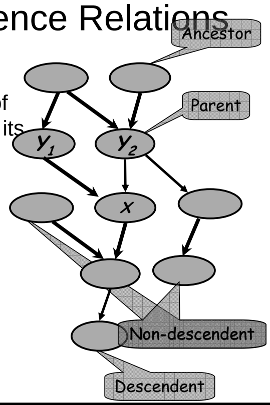


P(Lisa | Marge, Marge's parents) = P(Lisa | Marge).

Cond. Independence Relations

- Each random variable X, is conditionally independent of its non-descendents, given its parents:

  I(X; NonDesc(X) | Pa(X))



Independence

- The expression:

  $P(X_1,X_2,\ldots,X_n) = \prod_{i=1,\ldots,n} P(X_i|Pa(X_i))$

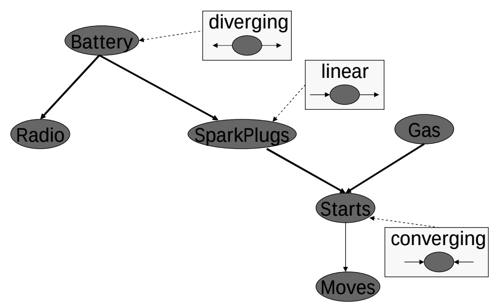
  means that each belief is independent of its predecessors in the BN given its parents
- Said otherwise, the parents of a belief X_i are all the beliefs that "directly influence" X_i
- Usually (but not always) the parents of X_i are its causes and X_i is the effect of these causes

E.g., JohnCalls is influenced by Burglary, but not directly. JohnCalls is directly influenced by Alarm

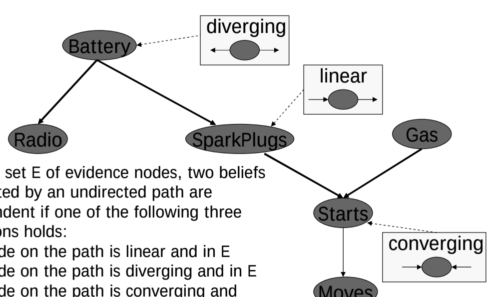
Example: Naïve Bayes Model

- A common model in disease diagnosis:
  - Symptoms are conditionally independent given the disease
- Thus, if
  - $X_1,\ldots,X_n$ denote whether the symptoms exhibited by the patient (headache, high-fever, etc.)
  - H denotes the hypothesis about the patient's health
- then, $P(X_1,\ldots,X_n,H) = P(H)P(X_1|H)\ldots P(X_n|H)$,
- This **naïve Bayes** model allows compact representation
  - It makes strong independence assumptions

Types of Nodes on a Path

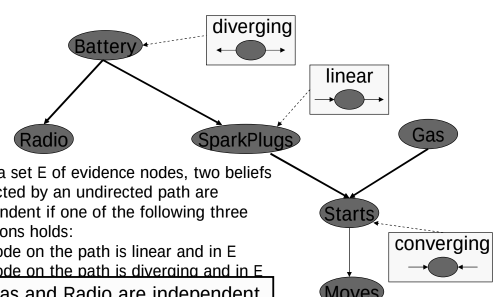


Independence Relations in BN



Given a set E of evidence nodes, two beliefs connected by an undirected path are independent if one of the following three conditions holds:
1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E

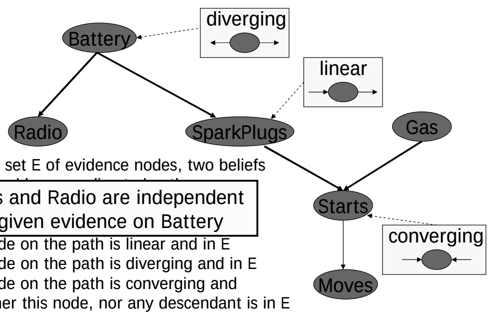
Independence Relations in BN



Given a set E of evidence nodes, two beliefs connected by an undirected path are independent if one of the following three conditions holds:
1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A [...] n E

Gas and Radio are independent given evidence on SparkPlugs

Independence Relations in BN

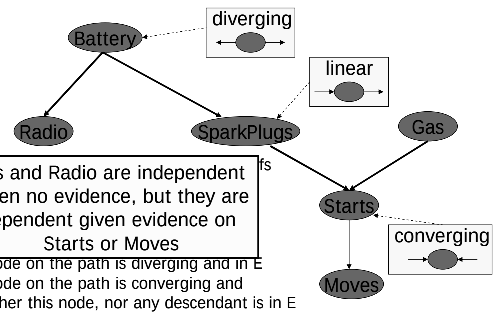


Given a set E of evidence nodes, two beliefs
co
ind
co

Gas and Radio are independent given evidence on Battery

1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E

Independence Relations in BN



G ... fs
co
in
co
1.

Gas and Radio are independent given no evidence, but they are dependent given evidence on Starts or Moves

2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E

DAGs and Topological Ordering

- Lemma: a directed graph is a DAG iff it has a topological ordering
- A topological ordering of a graph is an ordering of its nodes such that each node comes before all nodes to which it has edges (or in other words, once a n

How is this relevant to Bayesian networks?

- Proof: every DAG has a “sink”. Use it as the last node in the sort. Remove it from the DAG. The resulting graph is still a DAG. Proceed by induction. For the other direction assume the graph has a topological order, and yet is not a DAG. Existence of a cycle contradicts topological order.

Bayesian Networks

- A simple, graphical notation for conditional independence assertions resulting in a compact representation for the full joint distribution
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link = ‘direct influences’)
  - a conditional distribution for each node given its parents: $P(X_i|Parents(X_i))$