CS 188: Artificial Intelligence
Spring 2006

Lecture 15: Bayes' Nets

3/9/2006

Dan Klein – UC Berkeley

Outline

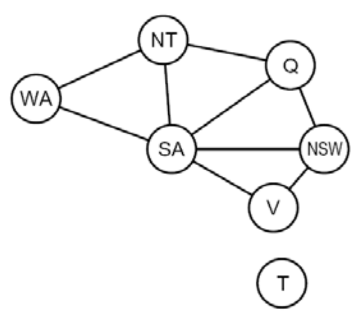
- Rest of course:
  - Bayes Nets
  - Speech Recognition / HMMs
  - Reinforcement learning
  - Applications: NLP, Vision, Games
- Today:
  - Bayes Nets Introduction

Models

- Models are descriptions of how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
- Why worry about probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

Reminder: CSPs

- CSPs were a kind of model
  - Describe legal interactions between variables
  - Usually we just look for some legal assignment
  - But, also can reason using all assignments, or find assignments consistent with evidence
- Key idea of CSPs:
  - Model global behavior using local constraints
  - Recurring idea in AI: compact local models interact to give efficient, interesting global behavior



$$D = \{red, green, blue\}$$

$$WA \neq NT$$

Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables

$$P(X_1, X_2, \ldots X_n)$$

- Given a joint distribution, we can reason about unobserved variables given observations (evidence)
- General form of a query:

$$P(x_q|x_{e_1}, \ldots x_{e_k})$$

Stuff you care about

Stuff you already know

- This kind of posterior distribution is also called the belief function of an agent which uses this model

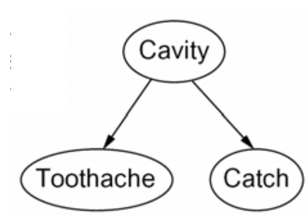
Bayes' Nets: Big Picture

- Two problems with generic probabilistic models:
  - Unless there are only a few variables, the joint is too big to represent explicitly
  - Hard to estimate anything empirically about more than a few variables at a time
- Bayes' nets are a technique for describing complex joint distributions (models) using a bunch of simple, local distributions
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we'll be very vague about how these interactions are specified

Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

- Arcs: interactions
  - Similar to constraints
  - Indicate "direct influence" between variables
- For now: imagine that arrows mean causation



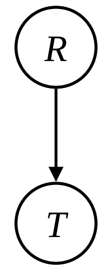
Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic
- Model 1: independence
- Model 2: rain causes traffic
- Why is an agent using model 2 better?



Example: Traffic II

- Let’s build a causal graphical model
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

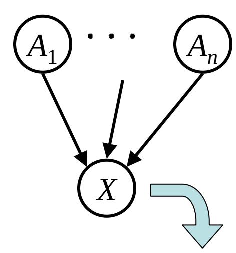
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!

Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A distribution over X, for each combination of parents' values

$$P(X|a_1 \ldots a_n)$$

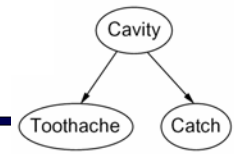
  - CPT: conditional probability table
  - Description of a noisy "causal" process



$$P(X|A_1 \ldots A_n)$$

A Bayes net = Topology (graph) + Local Conditional Probabilities

Probabilities in BNs



- Bayes' nets **implicitly** encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \textit{parents}(X_i))$$

  - Example:

$$P(\textit{cavity, catch}, \neg \textit{toothache})$$

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every full joint
  - The topology enforces certain conditional independencies

Example: Coin Flips

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

. . .

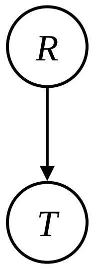
$P(X_n)$

h	0.5
t	0.5

$$P(h, h, t, h) =$$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic



$P(R)$

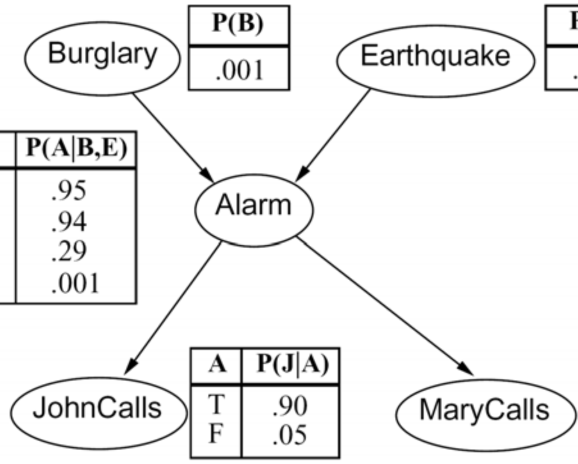
r	1/4
¬r	3/4

$P(T|R)$

r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(r, \neg t) =$

Example: Alarm Network



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

$P(b, e, \neg a, j, m) =$

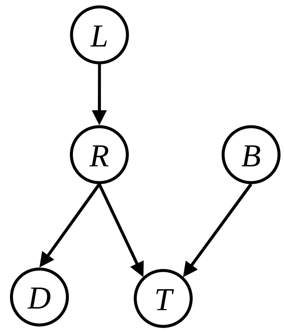
Example: Naïve Bayes

- Let's figure out what the Bayes' net for naïve Bayes is:

$$P(y, x_1, x_2 \dots x_n) = P(y)P(x_1|y)P(x_2|y)\dots P(x_n|y)$$

Example: Traffic II

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame



Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?
- How big is a Bayes net if each node has k parents?
- Both give you the power to calculate $P(X_1, X_2, \ldots X_n)$
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (next class)

Building the (Entire) Joint

- We can take a Bayes' net and build the full joint distribution it encodes

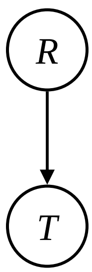
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \mathit{parents}(X_i))$$

  - Typically, there's no reason to do this
  - But it's important to know you could!

- To emphasize: every BN over a domain implicitly represents some joint distribution over that domain

Example: Traffic

- Basic traffic net
- Let's multiply out the joint



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

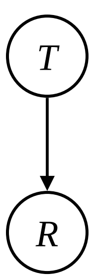
r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(T,R)$

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Example: Reverse Traffic

- Reverse causality?



$P(T)$

t	9/16
¬t	7/16

$P(R|T)$

t	r	1/3
	¬r	2/3
¬t	r	1/7
	¬r	6/7

$P(T,R)$

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independencies

Creating Bayes' Nets

- So far, we talked about how any fixed Bayes' net encodes a joint distribution
- Next: how to represent a fixed distribution as a Bayes' net
  - Key ingredient: conditional independence
  - The exercise we did in "causal" assembly of BNs was a kind of intuitive use of conditional independence
  - Now we have to formalize the process
- After that: how to answer queries (inference)