CS 188: Artificial Intelligence
Spring 2010

Lecture 15: Bayes' Nets II – Independence
3/9/2010

Pieter Abbeel – UC Berkeley

Many slides over the course adapted from Dan Klein, Stuart Russell, Andrew Moore

Announcements

- Current readings
  - Require login
- Assignments
  - W4 due Thursday
- Midterm
  - 3/18, 6-9pm, 0010 Evans --- no lecture on 3/18
  - We will be posting practice midterms
  - One page note sheet, non-programmable calculators
  - Topics go through Thursday, not next Tuesday

Outline

- Thus far: Probability
- Today: Bayes nets
  - Semantics
  - (Conditional) Independence

Probability recap

- Conditional probability $P(x|y) = \frac{P(x,y)}{P(y)}$
- Product rule $P(x,y) = P(x|y)P(y)$

always true

- Chain rule $P(X_1, X_2, \dots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\dots$
- X, Y independent iff: $\forall x, y : P(x,y) = P(x)P(y)$

$\Leftrightarrow P(x|y) = P(x) \Leftrightarrow P(y|x) = P(y)$

- X and Y are conditionally independent given Z iff:

$\forall x, y, z : P(x,y|z) = P(x|z)P(y|z)$ $\quad X \perp\!\!\!\perp Y | Z$

Bayes' Nets: Big Picture

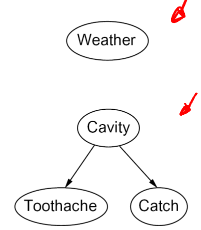
- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models

directed graphical models

  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we'll be 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 CSP constraints
  - Indicate "direct influence" between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don't!)



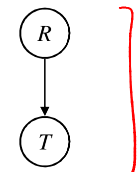
Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

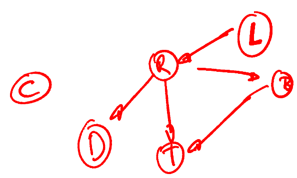
Example: Traffic

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



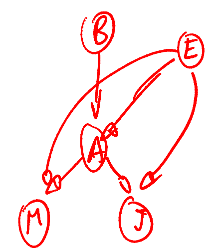
Example: Traffic II

- 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!

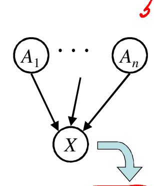


Bayes' Net Semantics

- 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 collection of distributions over X, one for each combination of parents' values

    $P(X|a_1 \dots a_n)$

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

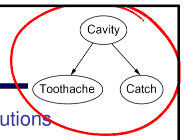


$P(X|A_1 \dots A_n)$

can think of it this way for now

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:

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

  - Example:

Chain rule: $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} P(x_i | x_1, \dots, x_{i-1})$

$P(+cavity,\ +catch,\ \neg toothache)$

$= P(+cavity) P(+catch|+cavity) * P(\neg toothache|+cavity)$

Conditional Indep.

Assumption $P(x_i | x_1, x_2, \dots x_{i-1}) = P(x_i | Parents(X_i))$

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

Example: Coin Flips

X_1 X_2 $\cdots$ X_n

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$\cdots$

$P(X_n)$

h	0.5
t	0.5

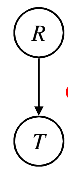
$P(h,h,t,h) = P(X_1=h)\cdot P(X_2=h)\cdot P(X_3=t)\cdot P(X_4=h)$

$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$

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

15

Example: Traffic



CPT $P(R)$

+r	1/4
¬r	3/4

CPT $P(T|R)$

+r	+t	3/4
	¬t	1/4
¬r	+t	1/2
	¬t	1/2

$P(+r,\neg t) = P(+r)$

$\cdot P(\neg t|+r)$

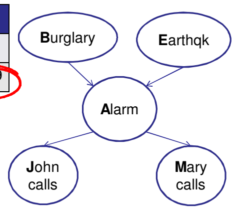
$= \frac{1}{4} * \frac{1}{4}$

$= \frac{1}{16}$

16

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

Example: Alarm Network



B	P(B)
+b	0.001
¬b	0.999

E	P(E)
+e	0.002
¬e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	¬a	0.05
+b	¬e	+a	0.94
+b	¬e	¬a	0.06
¬b	+e	+a	0.29
¬b	+e	¬a	0.71
¬b	¬e	+a	0.001
¬b	¬e	¬a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	¬j	0.1
¬a	+j	0.05
¬a	¬j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	¬m	0.3
¬a	+m	0.01
¬a	¬m	0.99

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?
  2^N -1 $2^{100} \approx 10^{30}$
- How big is an N-node net if nodes have up to k parents?
  O(N * 2^{k+1}) $N * 2^{k+1}$ $(N\ 2^k)$ $k=10$
- 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 (coming)

18

Bayes' Nets

- So far: how a Bayes' net encodes a joint distribution
- Next: how to answer queries about that distribution
  - Key idea: conditional independence
- After that: how to answer numerical queries (inference) more efficiently than by first constructing the joint distribution

19

Conditional Independence

- Reminder: independence
  - X and Y are independent if

$$\forall x,y\ \ P(x,y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

  - X and Y are conditionally independent given Z

$$\forall x,y,z\ \ P(x,y|z) = P(x|z)P(y|z) \dashrightarrow X \perp\!\!\!\perp Y | Z$$

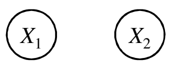
  - (Conditional) independence is a property of a distribution

$P(x_i \mid x_1, \ldots, x_{i-1}) = P(x_i \mid Pa(X_i))$

23

Example: Independence

- For this graph, you can fiddle with θ (the CPTs) all you want, but you won't be able to represent any distribution in which the flips are dependent!

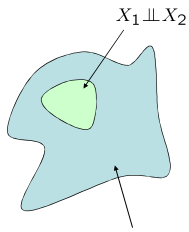


$P(X_1)$

h	0.5
t	0.5

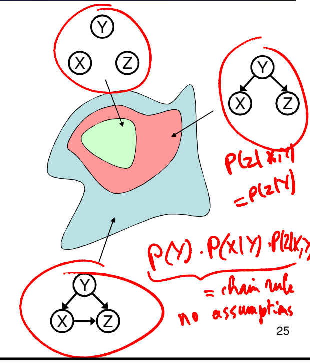
$P(X_2)$

h	0.5
t	0.5



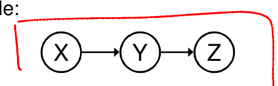
Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution



Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:

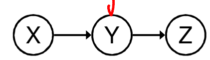


P(X) P(Y|X) P(Z|Y)

  - Question: are X and Z necessarily independent?
    - Answer: no. Example: low pressure causes rain, which causes traffic.
    - X can influence Z, Z can influence X (via Y)
    - Addendum: they *could* be independent: how?

Causal Chains

- This configuration is a "causal chain"



X: Low pressure
Y: Rain
Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is X independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

$$= P(z|y)$$

P(x,z|y) ≟ P(x|y) P(z|y)

P(x|y,z) ≟ P(x|y)

P(z|x,y) ≟ P(z|y) — Yes!

- Evidence along the chain "blocks" the influence

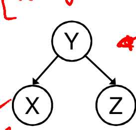
Common Cause

P(Y=y)=.5
P(x|y)=1 for x=y
P(z|y)=1 for z=y

- Another basic configuration: two effects of the same cause
  - Are X and Z independent? NO

Check by counterexample.

  - Are X and Z independent given Y?



Y: Project due
X: Newsgroup busy
Z: Lab full

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$

≟ P(z|y)

P(X,Y,Z) = P(Y)·P(X|Y) P(Z|Y)

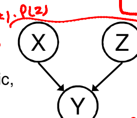
  - Observing the cause blocks influence between effects.

Causal chain: O→O→O, Common cause

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)

P(x,z) ≟ P(x)·P(z)
P(x|z) ≟ P(x)
P(z|x) ≟ P(z)



Y=xor(x,z)

X: Raining
Z: Ballgame
Y: Traffic

  - Are X and Z independent given Y?
    - No: seeing traffic puts the rain and the ballgame in competition as explanation?
  - This is backwards from the other cases
    - Observing an effect activates influence between possible causes.

The General Case

- Any complex example can be analyzed using these three canonical cases
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph

30