Bayes Nets 2

Ch. 14.1-2,4

Adapted from slides kindly shared by Stuart Russell

Appreciations

- ◊ Thanksgiving!!
- ◊ Roe McBurnett Jr and the Power of Positive Thinking

Share some of yours?

Announcements

Project P3 Reinforcement due Thu Nov 29th at 17:00

Homework on Bayes Nets: “Green Party President”

Solution at “Green Party President solution”

Outline

◇ Bayes Nets, D-separation

Credit to Pieter Abbeel, Dan Klein, Stuart Russell and Andrew Moore for most of today's slides

Probability recap

- Conditional probability $P(x|y) = \dfrac{P(x,y)}{P(y)}$
- Product rule $P(x,y) = P(x|y)P(y)$
- Chain rule
$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1}) \end{aligned}$$
- X, Y independent iff: $\forall x,y : P(x,y) = P(x)P(y)$
- X and Y are conditionally independent given Z iff:

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

Bayes' Nets

- Representation
  - Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets
- Conditional Independences
- Probabilistic Inference
- Learning Bayes' Nets from Data

Build your own Bayes nets!

- http://www.aispace.org/bayes/index.shtml

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

  $O(N * 2^{k+1})$

- 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)

Bayes' Nets

- Representation
  - Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets
- Conditional Independences
- Probabilistic Inference
- Learning Bayes' Nets from Data

Representing Joint Probability Distributions

- **Table representation**:

  number of parameters: d^n-1

- **Chain rule representation:**

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

number of parameters: (d-1) + d(d-1) + d^2(d-1)+…+d^{n-1}(d-1) = d^n-1

Size of CPT = (number of different joint instantiations of the preceding variables) *times* (number of values current variable can take on *minus* 1)

- Both can represent any distribution over the n random variables. Makes sense same number of parameters needs to be stored.
- Chain rule applies to all orderings of the variables, so for a given distribution we can represent it in n! = n factorial = n(n-1)(n-2)…2.1 different ways with the chain rule

Chain Rule → Bayes' net

- **Chain rule representation: applies to ALL distributions**
  - Pick any ordering of variables, rename accordingly as x_1, x_2, …, x_n

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i | x_1 \ldots x_{i-1})$$

Exponential in n

number of parameters: $(d-1) + d(d-1) + d^2(d-1) + \ldots + d^{n-1}(d-1) = d^n - 1$

- **Bayes' net representation: makes assumptions**
  - Pick any ordering of variables, rename accordingly as x_1, x_2, …, x_n
  - Pick any directed acyclic graph consistent with the ordering
  - *Assume* following conditional independencies:

$$P(x_i | x_1 \cdots x_{i-1}) = P(x_i | parents(X_i))$$

→ **Joint:** $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))$

number of parameters: (maximum number of parents = K)

$$\sum_{i=1}^{n} d^{|\text{parents}(X_i)|}(d-1) = O(nd^K(d-1)) = O(nd^{K+1})$$

Linear in n

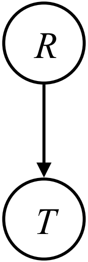
Note: no causality assumption made anywhere.

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 only guaranteed to encode conditional independence

Example: Traffic

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



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

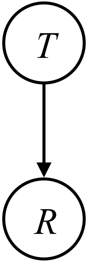
r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(T, R)$

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

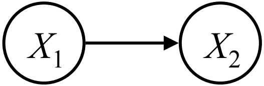
¬t	r	1/7
	¬r	6/7

$P(T, R)$

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

Example: Coins

- Extra arcs don't prevent representing independence, just allow non-independence



$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$P(X_1)$

h	0.5
t	0.5

$P(X_2|X_1)$

h | h	0.5
t | h	0.5

h | t	0.5
t | t	0.5

- Adding unneeded arcs isn't wrong, it's just inefficient

Bayes' Nets

- Representation
  - ✓ Informal first introduction of Bayes' nets through causality "intuition"
  - ✓ More formal introduction of Bayes' nets
- Conditional Independences
- Probabilistic Inference
- Learning Bayes' Nets from Data

Bayes Nets: Assumptions

- To go from chain rule to Bayes' net representation, we made the following assumption about the distribution:

$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Turns out that probability distributions that satisfy the above ("chain-rule→Bayes net") conditional independence assumptions
  - often can be guaranteed to have many more conditional independences
  - These guaranteed additional conditional independences can be read off directly from the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph

Example

$$X \rightarrow Y \rightarrow Z \rightarrow W$$

- Conditional independence assumptions directly from simplifications in chain rule:

- Additional implied conditional independence assumptions?

Independence in a BN

- Given a Bayes net graph
  - Important question:

    Are two nodes guaranteed to be independent given certain evidence?

  Equivalent question:

    Are two nodes independent given the evidence in all distributions that can be encoded with the Bayes net graph?

- Before proceeding: How about opposite question: Are two nodes guaranteed to be *dependent* given certain evidence?
  - No! For any BN graph you can choose all CPT's such that all variables are independent by having P(X | Pa(X) = paX) not depend on the value of the parents. Simple way of doing so: pick all entries in all CPTs equal to 0.5 (assuming binary variables)

Independence in a BN

- Given a Bayes net graph

  Are two nodes guaranteed to be independent given certain evidence?

  - ***If no***, can prove with a counter example
    - I.e., pick a distribution that can be encoded with the BN graph, i.e., pick a set of CPT's, and show that the independence assumption is violated
  - ***If yes***,
    - For now we are able to prove using algebra (tedious in general)
    - Next we will study an efficient graph-based method to prove yes: "D-separation"

D-separation: Outline

- Study independence properties for triples
- Any complex example can be analyzed by considering relevant triples

Causal Chains

- This configuration is a "causal chain"

$$X \rightarrow Y \rightarrow Z$$

X: Low pressure

Y: Rain

Z: Traffic

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

- Is it guaranteed that X is independent of Z ? *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example: P(y|x) = 1 if y=x, 0 otherwise
    P(z|y) = 1 if z=y, 0 otherwise
    Then we have P(z|x) = 1 if z=x, 0 otherwise
    hence X and Z are not independent in this example

Causal Chains

- This configuration is a "causal chain"

$$X \rightarrow Y \rightarrow Z$$

X: Low pressure

Y: Rain

Z: Traffic

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

- Is it guaranteed that X is independent of Z given Y?

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

$$= P(z|y)$$

Yes!

- Evidence along the chain "blocks" the influence

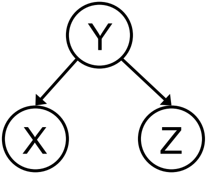
Common Cause

- Another basic configuration: two effects of the same cause
  - Is it guaranteed that X and Z are independent?
    - *No!*
    - Counterexample:

Choose P(X|Y)=1 if x=y, 0 otherwise,

Choose P(z|y) = 1 if z=y, 0 otherwise.

Then P(x|z)=1 if x=z and 0 otherwise, hence X and Z are not independent in this example and hence it is not guaranteed that if a distribution can be encoded with the Bayes' net structure on the right that X and Z are independent in that distribution



Y: Project due

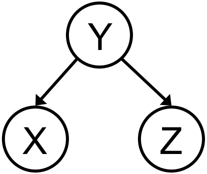
X: Piazza busy

Z: Lab full

Common Cause

- Another basic configuration: two effects of the same cause

  - Is it guaranteed that X and Z are independent given Y?



Y: Project due

X: Piazza busy

Z: Lab full

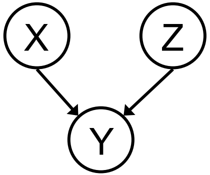
$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$ *Yes!*

  - Observing the cause blocks influence between effects.

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!)
  - 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.



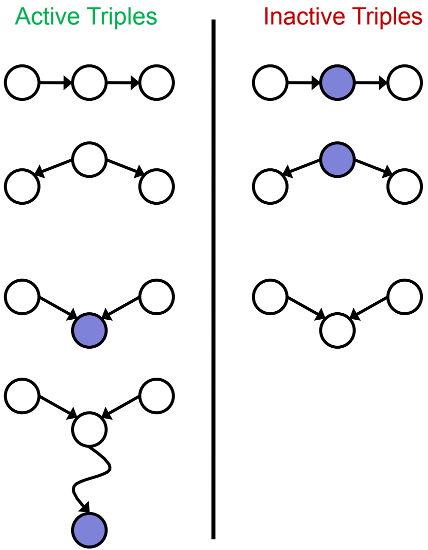
X: Raining

Z: Ballgame

Y: Traffic

Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Yes, if X and Y "separated" by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
  - Common effect (aka v-structure) $A \rightarrow B \leftarrow C$ where B *or one of its descendents* is observed
- All it takes to block a path is a single inactive segment



D-Separation

- Given query $X_i \overset{?}{\perp\!\!\!\perp} X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Shade all evidence nodes
- For all (undirected!) paths between and
  - Check whether path is active
    - If active return:

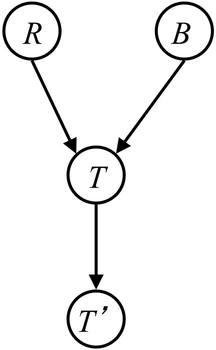
      not guaranteed that $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$
- (If reaching this point all paths have been checked and shown inactive)
  - Return: guaranteed tat $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$

Example

$R \perp\!\!\!\perp B$ Yes

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example

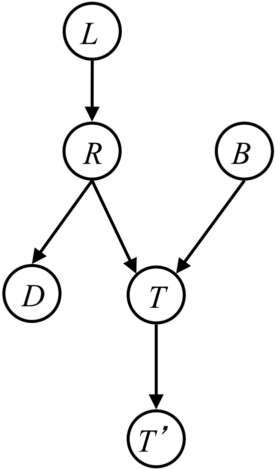
$L \perp\!\!\!\perp T' | T$ Yes

$L \perp\!\!\!\perp B$ Yes

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ Yes



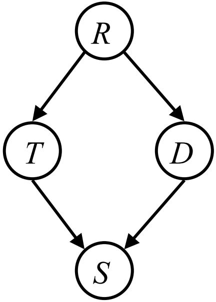
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I' m sad
- Questions:

$$T \perp\!\!\!\perp D$$

$$T \perp\!\!\!\perp D | R$$ *Yes*

$$T \perp\!\!\!\perp D | R, S$$



All Conditional Independences

- Given a Bayes net structure, can run d-separation to build a complete list of conditional independences that are guaranteed to be true, all of the form

$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

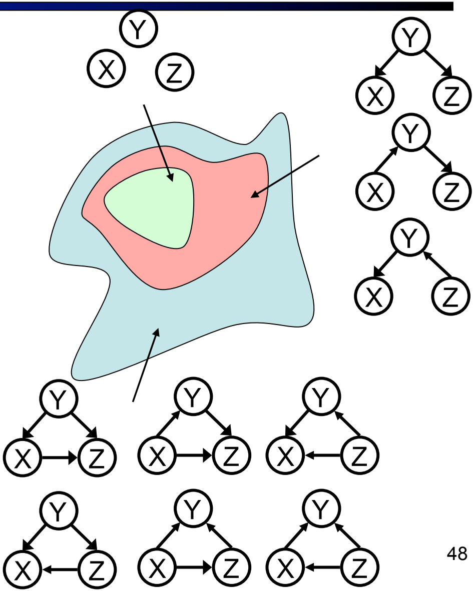
Possible to have same full list of conditional independencies for different BN graphs?

- Yes!
- Examples:

- If two Bayes' Net graphs have the same full list of conditional independencies then they are able to encode the same set of distributions.

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



Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution

Bayes' Nets

- ✓ Representation
- ✓ Conditional Independences
- Probabilistic Inference
  - Enumeration (exact, exponential complexity)
  - Variable elimination (exact, worst-case exponential complexity, often better)
  - Probabilistic inference is NP-complete
  - Sampling (approximate)
- Learning Bayes' Nets from Data