Bayes Net Learning

Ronald J. Williams
CSG220
Spring 2007

A very slightly modified version of
the Andrew Moore tutorial
Bayes Net Structure Learning

Reminder: A Bayes Net
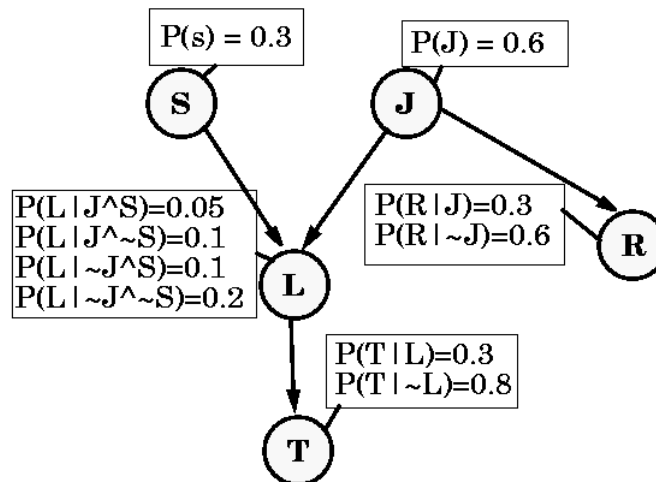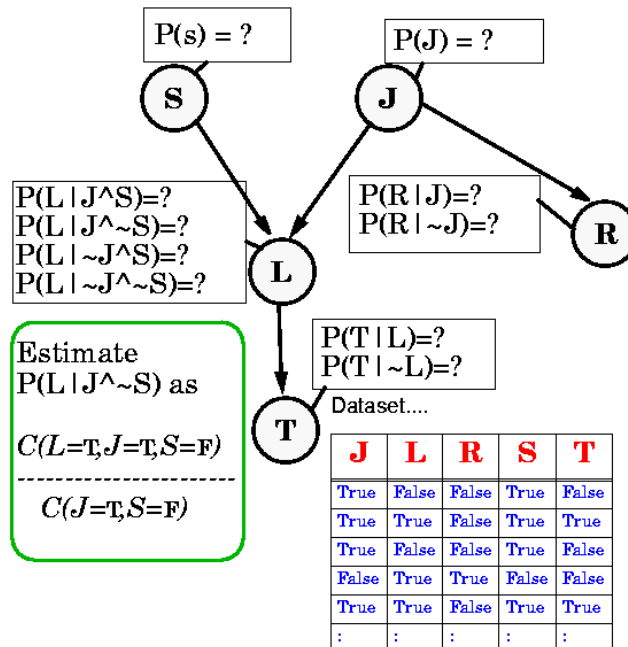
Two Forms of Bayes Net Learning from Data

- Given structure, determine probability tables
  - Straightforward
- Determine the structure as well as the probability tables
  - Much harder
  - Involves trying to estimate when conditional independence is or is not likely based on the data
  - Also requires estimating the probability tables
  - May also involve adding new nodes

Estimating Probability Tables

$P(s) = ?$

$P(J) = ?$

$P(L|J \wedge S)=?$
$P(L|J \wedge \sim S)=?$
$P(L|\sim J \wedge S)=?$
$P(L|\sim J \wedge \sim S)=?$

$P(R|J)=?$
$P(R|\sim J)=?$

$P(T|L)=?$
$P(T|\sim L)=?$

Estimate
$P(L|J \wedge \sim S)$ as

$$\frac{C(L=\text{T}, J=\text{T}, S=\text{F})}{C(J=\text{T}, S=\text{F})}$$

Dataset....

J	L	R	S	T
True	False	False	True	False
True	True	False	True	True
True	False	False	True	False
False	True	True	False	False
True	True	False	True	True
:	:	:	:	:

2

Slide 1: Estimating Probability Tables

Estimating Probability Tables

P(s) = ?

P(J) = ?

ct(J)

ct(S)

P(L|J^S)=?
P(L|J^~S)=?
P(L|~J^S)=?
P(L|~J^~S)=?

P(R|J)=?
P(R|~J)=?

ct(R,J)

ct(J,L,S)

P(T|L)=?
P(T|~L)=?

Dataset....

ct(L,T)

J	L	R	S	T
True	False	False	True	False
True	True	False	True	True
True	False	False	True	False
False	True	True	False	False
True	True	False	True	True
:	:	:	:	:
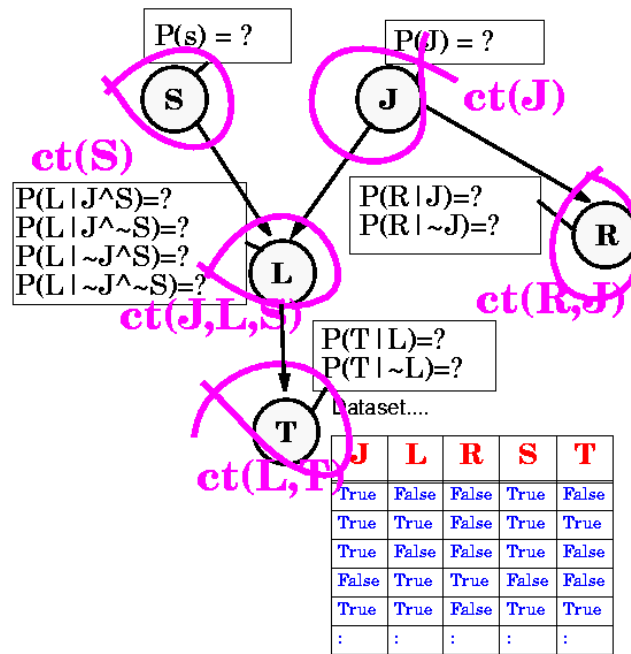
Bayes Net Learning: Slide 5

Slide 2: Scoring a structure

Scoring a structure

(Which of these fits the data best?)

N. Friedman and Z. Yakhini, On the sample complexity of learning Bayesian networks, Proceedings of the 12th conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1996

$$\text{Score} = -\frac{N_{\text{params}}}{2}\log R + R\sum_{j=1}^{m} \sum_{k=1}^{\left(\begin{array}{c}\text{num combinations}\\\text{of parent values}\end{array}\right)} \sum_{v=1}^{(\text{arity of } X_j)} P(V_k)P(X_j = v | V_k) \log P(X_j = v | V_k)$$

Bayes Net Learning: Slide 6

3

Scoring a structure

Number of non-redundant parameters defining the net

#Attributes

#Records

Sums over all the rows in the probability table for X_j

The parent values in the k'th row of X_j's probability table

$$\text{Score} = $$

$$-\frac{N_{\text{params}}}{2}\log R$$

$$+R\sum_{j=1}^{m}\overset{\left(\substack{\text{num combinations}\\\text{of parent values}}\right)}{\sum_{k=1}}\overset{(\text{arity of }X_j)}{\sum_{v=1}}P(V_k)P(X_j=v\,|\,V_k)\log P(X_j=v\,|\,V_k)$$

All these values estimated from data

Bayes Net Learning: Slide 7

Scoring a structure

This is called a BIC (Bayes Information Criterion) estimate
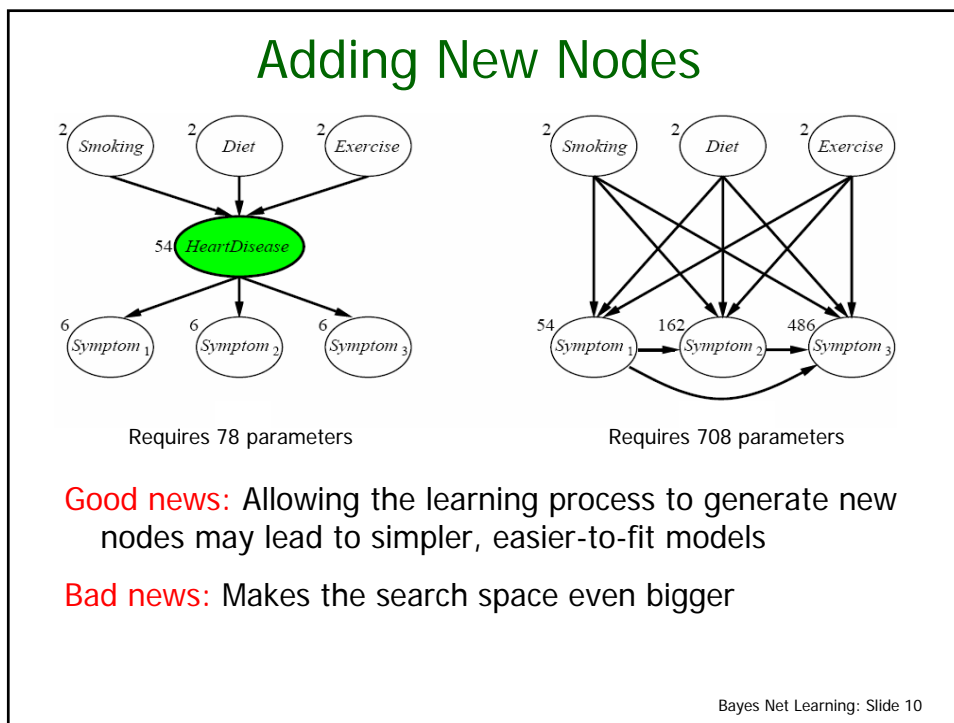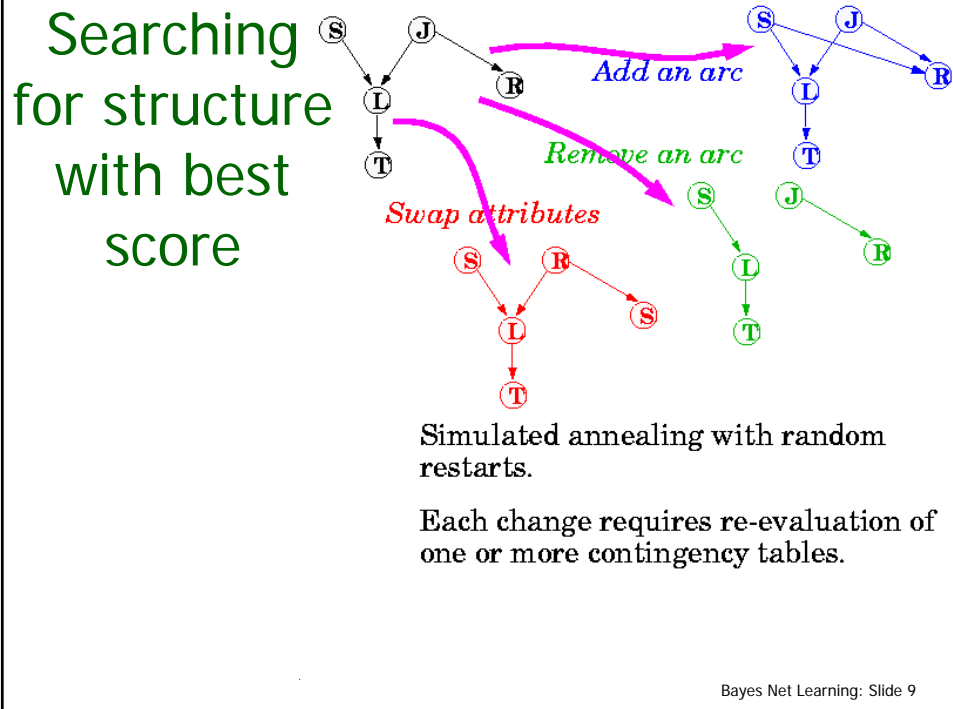
This part is a penalty for too many parameters

This part is the training set log-likelihood

BIC asymptotically tries to get the structure right. (There's a lot of heavy emotional debate about whether this is the best scoring criterion)

$$\text{Score} = $$

$$-\frac{N_{\text{params}}}{2}\log R$$

$$+R\sum_{j=1}^{m}\overset{\left(\substack{\text{num combinations}\\\text{of parent values}}\right)}{\sum_{k=1}}\overset{(\text{arity of }X_j)}{\sum_{v=1}}P(V_k)P(X_j=v\,|\,V_k)\log P(X_j=v\,|\,V_k)$$

All these values estimated from data

Bayes Net Learning: Slide 8

4

Searching for structure with best score

Add an arc

Remove an arc

Swap attributes

Simulated annealing with random restarts.

Each change requires re-evaluation of one or more contingency tables.

Adding New Nodes

Requires 78 parameters

Requires 708 parameters

Good news: Allowing the learning process to generate new nodes may lead to simpler, easier-to-fit models

Bad news: Makes the search space even bigger

What you should know

- That there exist methods for trying to learn Bayes nets from data