

Note to other teachers and users of these slides. Andrew would be delighted if you found this source material useful in giving your own lectures. Feel free to use these slides verbatim, or to modify them to fit your own needs. PowerPoint originals are available. If you make use of a significant portion of these slides in your own lecture, please include this message, or the following link to the source repository of Andrew's tutorials: <http://www.cs.cmu.edu/~awm/tutorials>. Comments and corrections gratefully received.

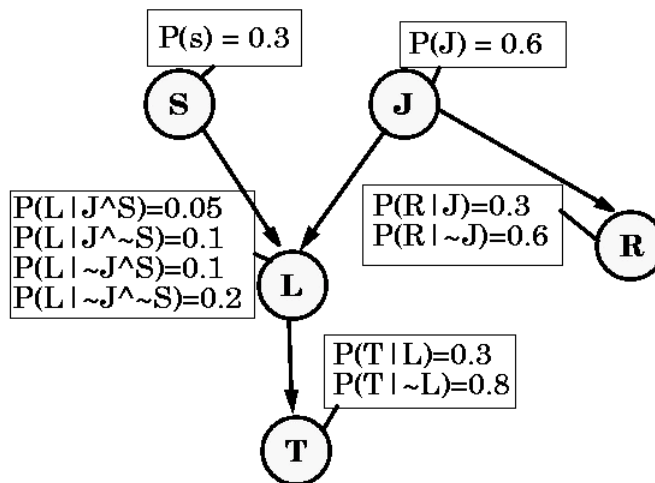
Bayes Net Structure Learning

Andrew W. Moore
Associate Professor
School of Computer Science
Carnegie Mellon University
www.cs.cmu.edu/~awm
awm@cs.cmu.edu
412-268-7599

Copyright © 2001, Andrew W. Moore

Oct 29th, 2001

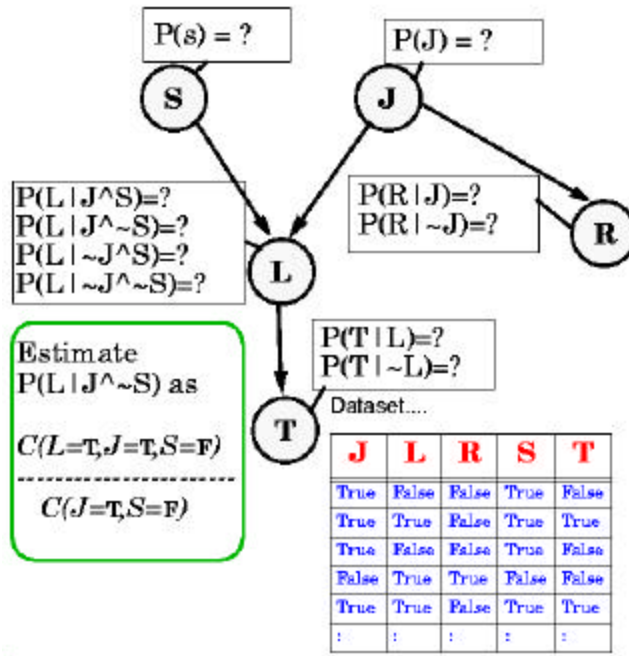
Reminder: A Bayes Net



Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 2

Estimating Probability Tables



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

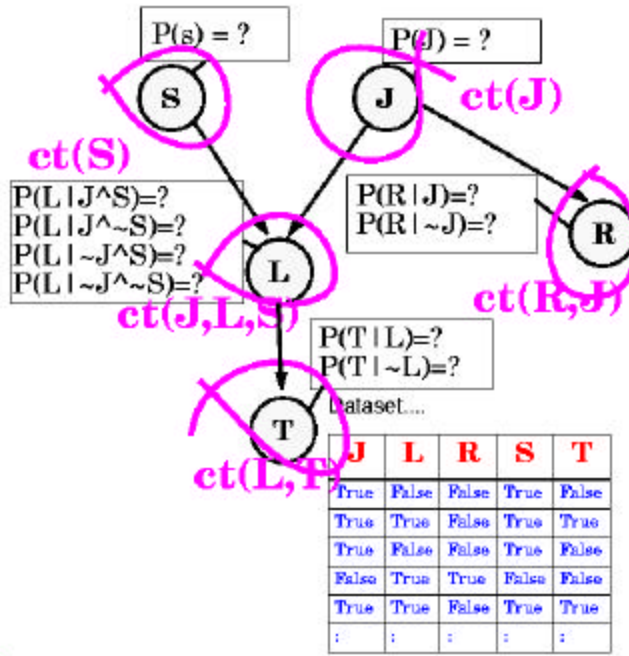
$C(L=T, J=T, S=F)$

$C(J=T, S=F)$

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 3

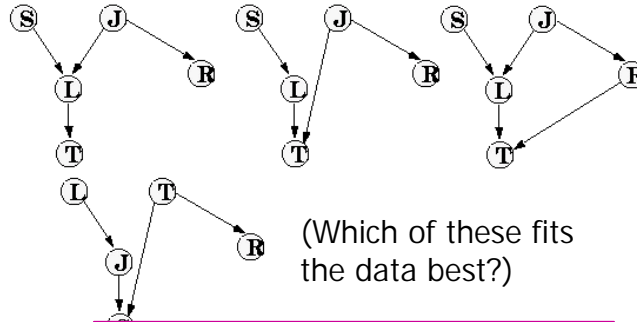
Estimating Probability Tables



Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 4

Scoring a structure

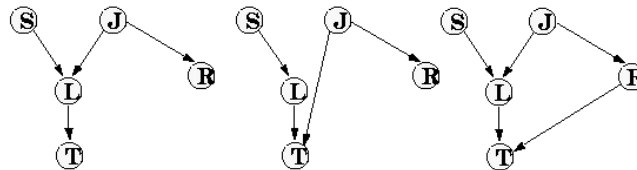


(Which of these fits the data best?)

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

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

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

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

All these values estimated from data

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

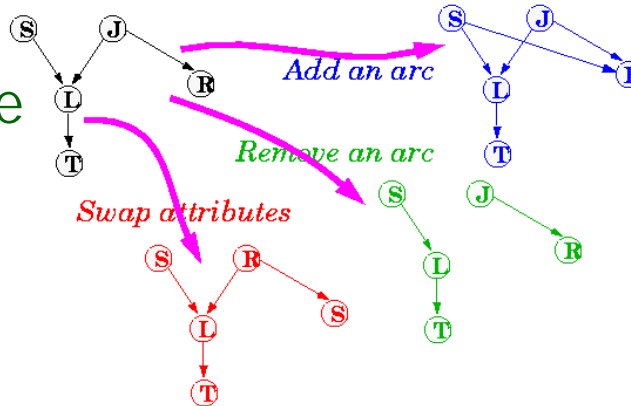
BIC asymptotically tries to get the structure right. (There's a lot of heavy emotional debate about whether this is the best scoring criterion)

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

$$+ R \sum_{j=1}^m \sum_{k=1}^{\left(\begin{smallmatrix} \text{num combinations} \\ \text{of parent vales} \end{smallmatrix}\right)} P(V_k) P(X_j = v | V_k) \log P(X_j = v | V_k)$$

All these values estimated from data



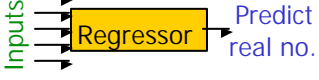
Searching for structure with best score



Simulated annealing with random restarts.

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


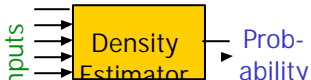
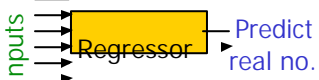
Learning Methods until today

	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh
	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE
	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 9

Learning Methods added today

	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh
	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE, Bayes Net Structure Learning (Note, can be extended to permit mixed categorical/real values)
	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 10

But also, for free...

	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh, Bayes Net Based BC
	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE, Bayes Net Structure Learning
	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 11

And a new operation...

	Joint DE, Bayes Net Structure Learning
	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh, Bayes Net Based BC
	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE, Bayes Net Structure Learning
	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 12