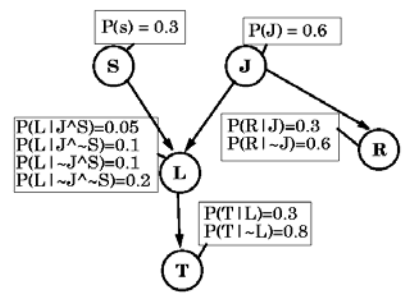
Bayes Net Structure Learning

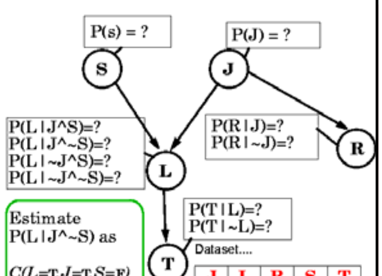
Andrew W. Moore
Associate Professor
School of Computer Science
Carnegie Mellon University
www.cs.cmu.edu/~awm
awm@cs.cmu.edu
412-268-7599

 Oct 29th, 2001

Reminder: A Bayes Net



Estimating Probability Tables



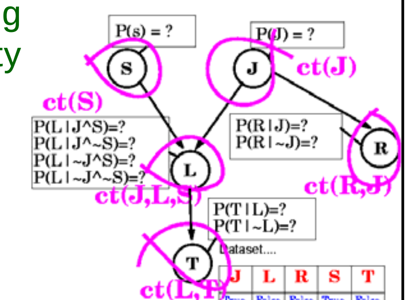
Estimate P(L | J^~S) as

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

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
:	:	:	:	:

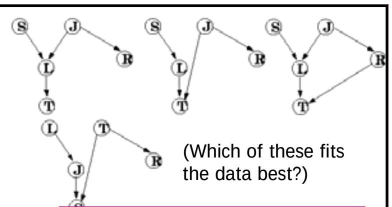
Estimating Probability Tables



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
:	:	:	:	:

Scoring a structure

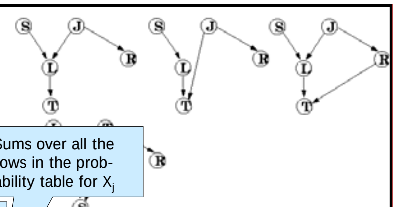


(Which of these fits the data best?)

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

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

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

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

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

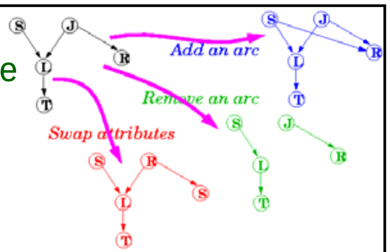
This part is the training set log-likelihood

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

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

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

Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh
Inputs → Density Estimator → Prob-ability	Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE
Inputs → Regressor → Predict real no.	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Learning Methods added today

Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh
Inputs → Density Estimator → Prob-ability	Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE, Bayes Net Structure Learning (Note, can be extended to permit mixed categorical/real values)
Inputs → Regressor → Predict real no.	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

But also, for free...

Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh, Bayes Net Based BC
Inputs → Density Estimator → Prob-ability	Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE, Bayes Net Structure Learning
Inputs → Regressor → Predict real no.	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

And a new operation...

Inputs → Inference Engine Learn → $P(E_1\|E_2)$	Joint DE, Bayes Net Structure Learning
Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh, Bayes Net Based BC
Inputs → Density Estimator → Prob-ability	Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE, Bayes Net Structure Learning
Inputs → Regressor → Predict real no.	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR