

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

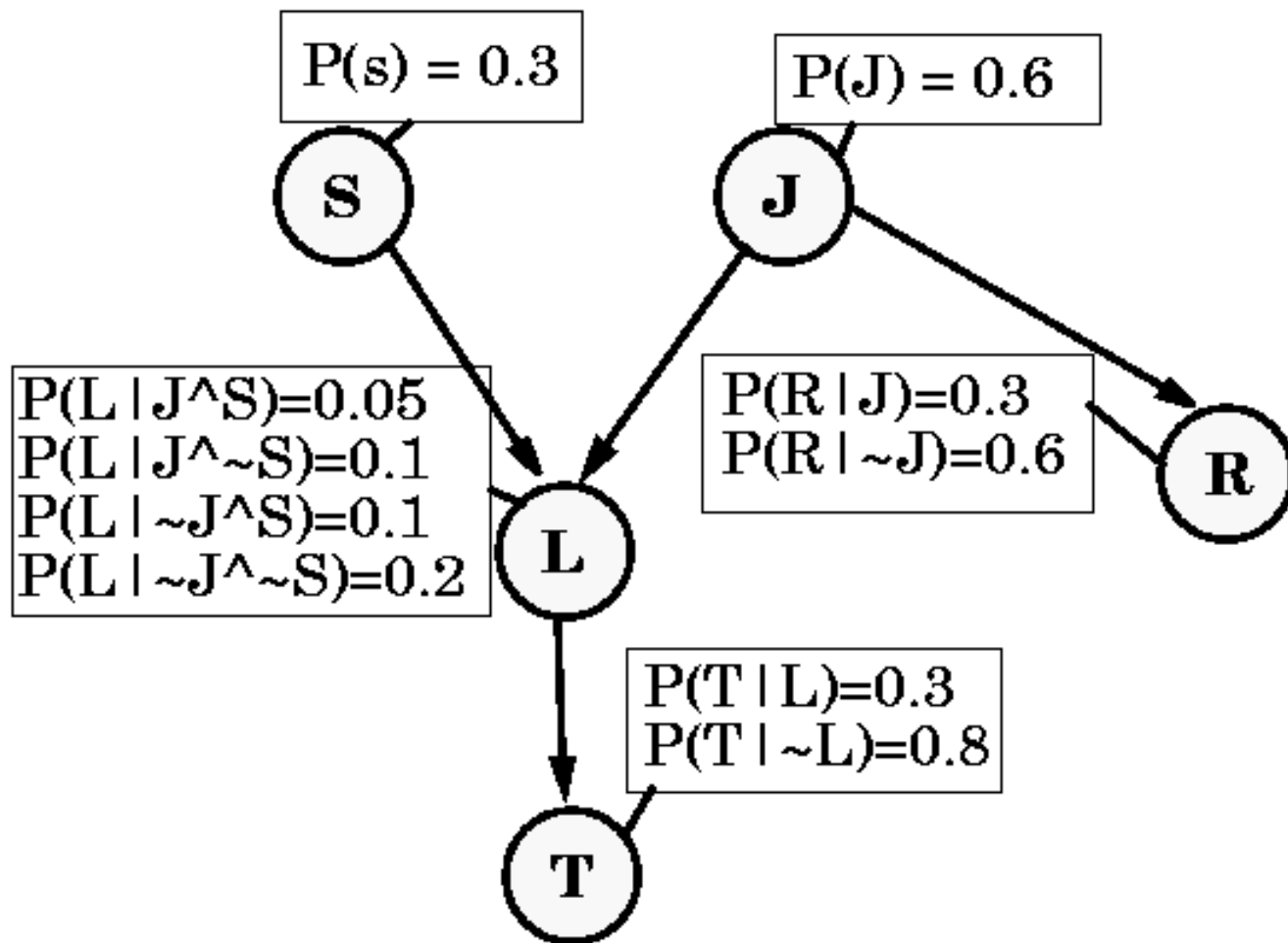
**Andrew W. Moore**  
**Associate Professor**  
**School of Computer Science**  
**Carnegie Mellon University**

[www.cs.cmu.edu/~awm](http://www.cs.cmu.edu/~awm)

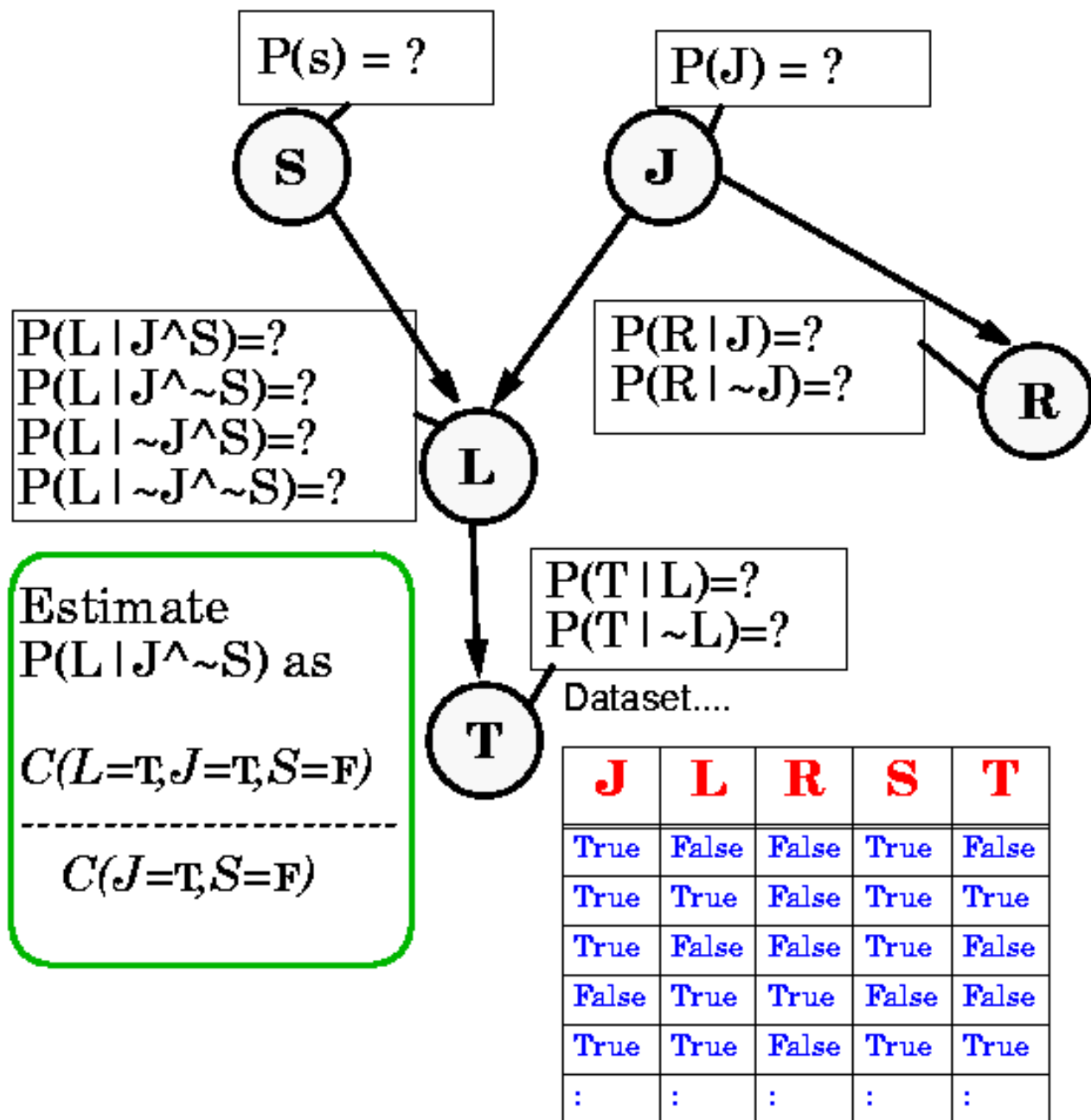
[awm@cs.cmu.edu](mailto:awm@cs.cmu.edu)

412-268-7599

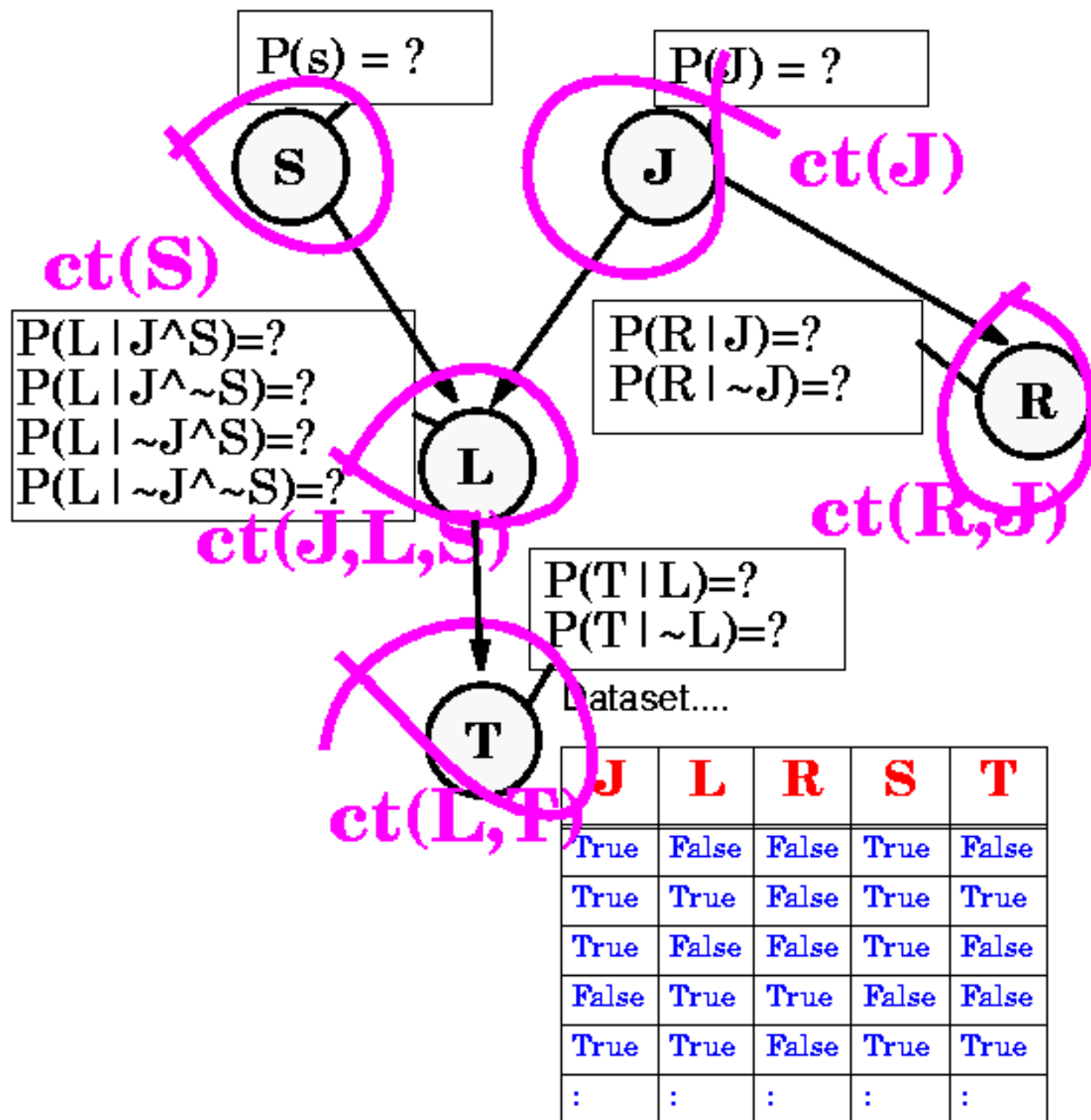
# Reminder: A Bayes Net



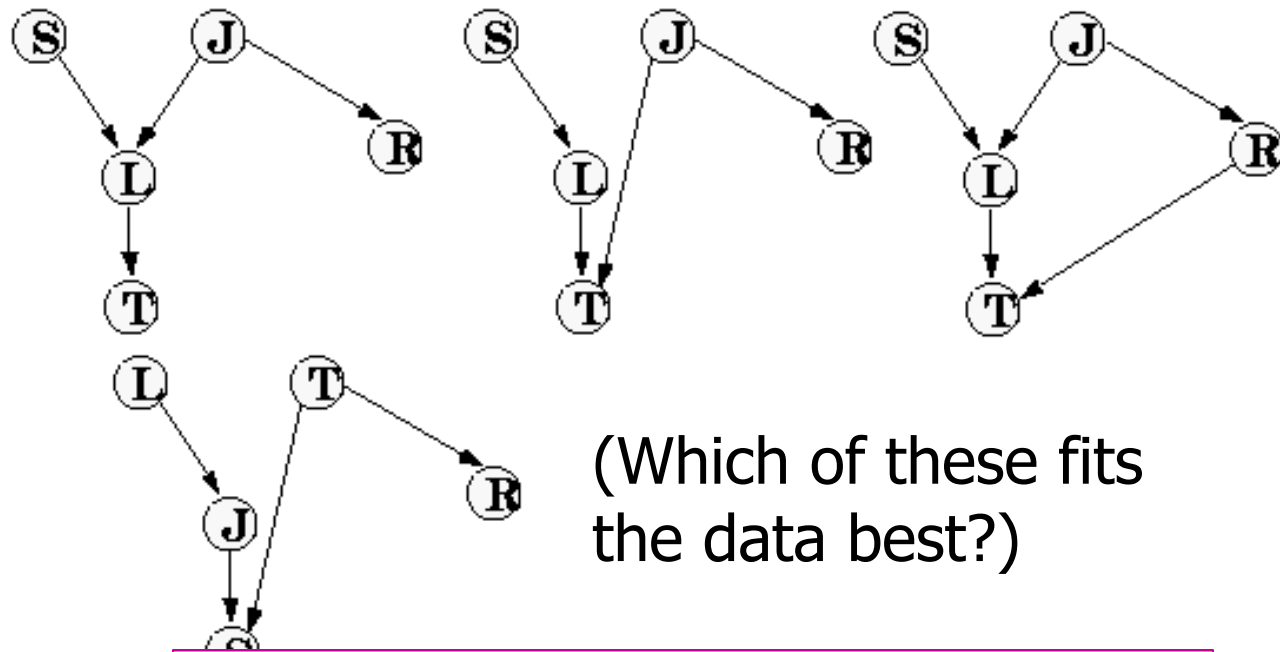
# Estimating Probability Tables



# Estimating Probability Tables



# Scoring a structure

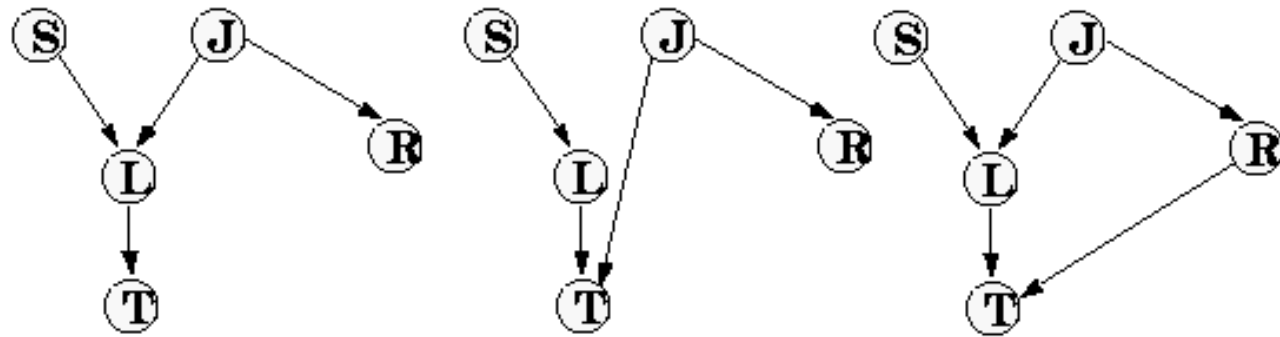


$$\text{Score} = -\frac{N_{\text{params}}}{2} \log R$$

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

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

Sums over all the rows in the probability table for  $X_j$

#Records

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

$$\begin{aligned}
 \text{Score} &= -\frac{N_{\text{params}}}{2} \log R \\
 &+ R \sum_{j=1}^m \sum_{k=1}^{\left( \begin{smallmatrix} \text{num combinations} \\ \text{of parent values} \end{smallmatrix} \right) (\text{arity of } X_j)} \sum_{v=1}^{\text{arity of } X_j} \underbrace{P(V_k)}_{\text{estimated from data}} \underbrace{P(X_j = v | V_k)}_{\text{estimated from data}} \log \underbrace{P(X_j = v | V_k)}_{\text{estimated from data}}
 \end{aligned}$$

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

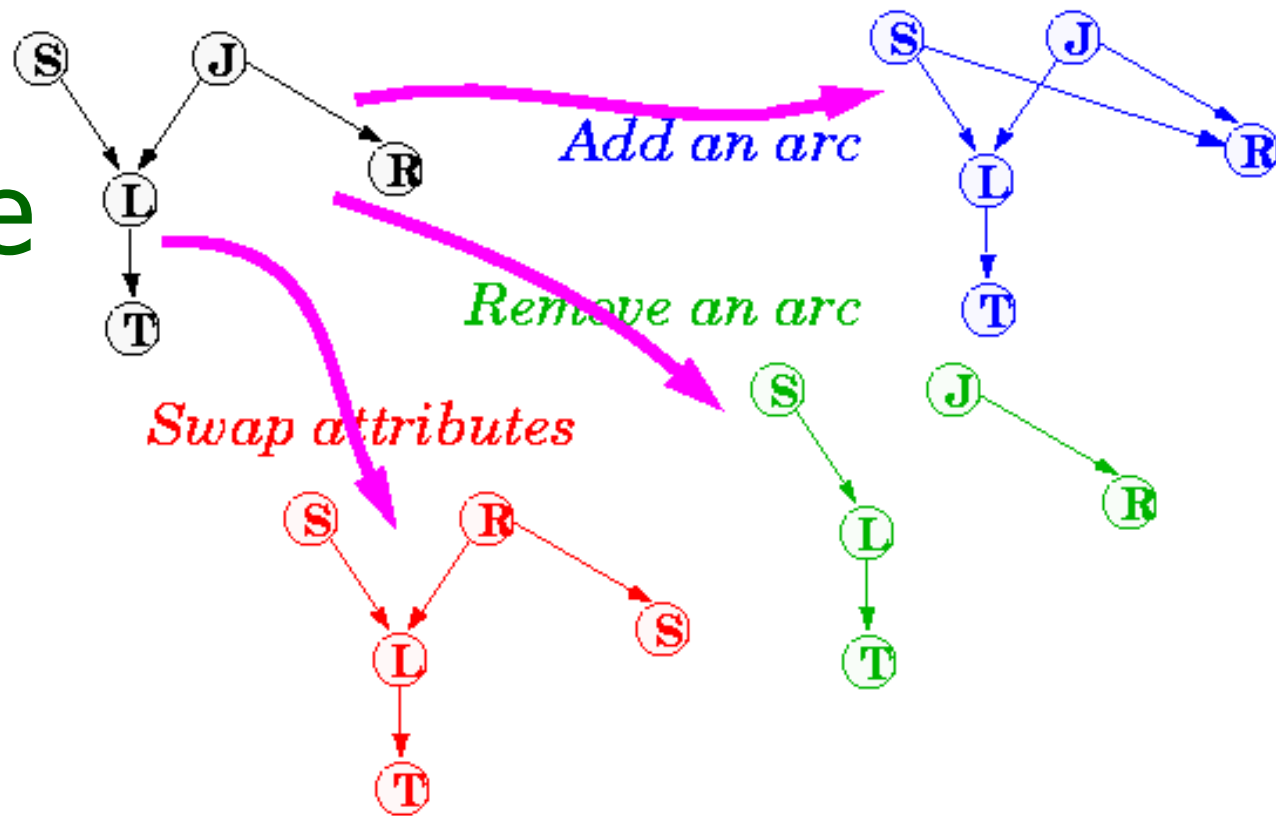
BIC asymptotically tries to get the structure right. (There's a lot of heavy emotional debate about whether this is the best scoring criterion)

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

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

All these values estimated from data

# 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, <b>Bayes Net Structure Learning (Note, can be extended to permit mixed categorical/real values)</b>
Inputs — — — — —	Regressor	Predict real no.	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

# But also, for free...

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

# 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	Probability	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