Oct 29th, 2001

Copyright © 2001, Andrew W. Moore

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 .](http://www.cs.cmu.edu/%7Eawm/tutorials)

Comments and corrections gratefully received.

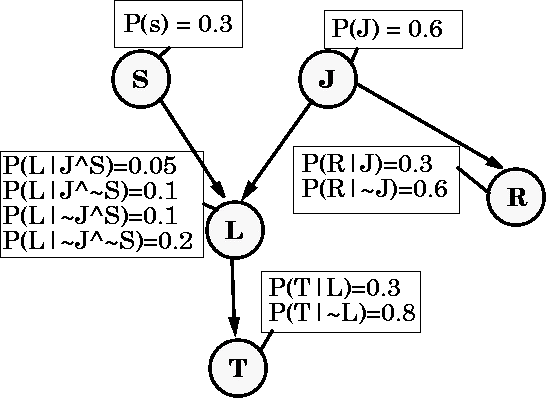
**Bayes Net Structure Learning**

**Andrew W. Moore Associate Professor**

**School of Computer Science Carnegie Mellon University**

[www.cs.cmu.edu/~awm](http://www.cs.cmu.edu/%7Eawm)

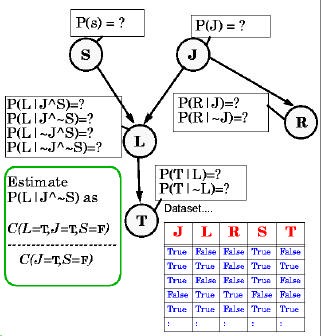
[awm@cs.cmu.edu](mailto:awm@cs.cmu.edu) 412-268-7599



Bayes Net Structure: Slide 2

Copyright © 2001, Andrew W. Moore

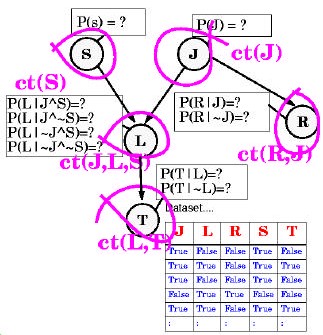
Reminder: A Bayes Net



Estimating Probability Tables

Copyright © 2001, Andrew W. Moore

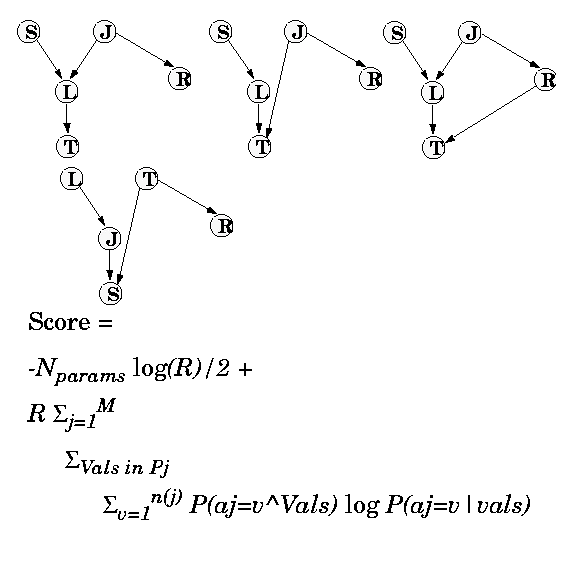
Bayes Net Structure: Slide 3



Estimating Probability Tables

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 4



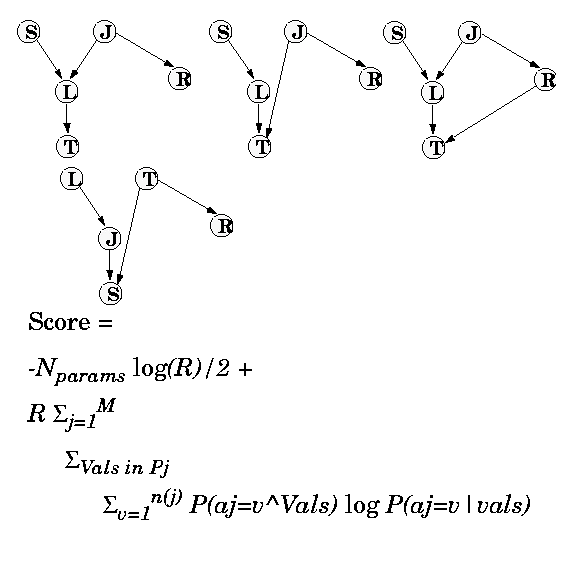
Scoring a structure

(Which of these fits the data best?)

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 5

|  |  |  |
| --- | --- | --- |
|  | N. Friedman and Z. Yakhini, On the sample |  |
| Score    * *N*params log *R*   2 | complexity of learning Bayesian networks, Proceedings of the 12th conference on Uncertainty in Artificial Intelligence, Morgan  Kaufmann, 1996 |  |
|  num combinations  *m*  of parent vaules  (arityof *X j* )   * *R*   *P*(*Vk* )*P*( *X j*  *v* | *Vk* ) log *P*( *X j*  *v* |*Vk* )   *j* 1 *k* 1 *v* 1 | | |



 *P*(*Vk* )*P*( *X j*  *v* | *Vk* ) log *P*( *X j*  *v* |*Vk* )

*v* 1

All these values estimated from data

Bayes Net Structure: Slide 6

Copyright © 2001, Andrew W. Moore



*k* 1

*m*  of parent vaules  (arityof *X j* )

* *R*

*j* 1

 num combinations

in the k’th row of Xj’s probability table

2

* *N*params log *R*

Sums over all the rows in the prob- ability table for Xj

#Records

The parent values

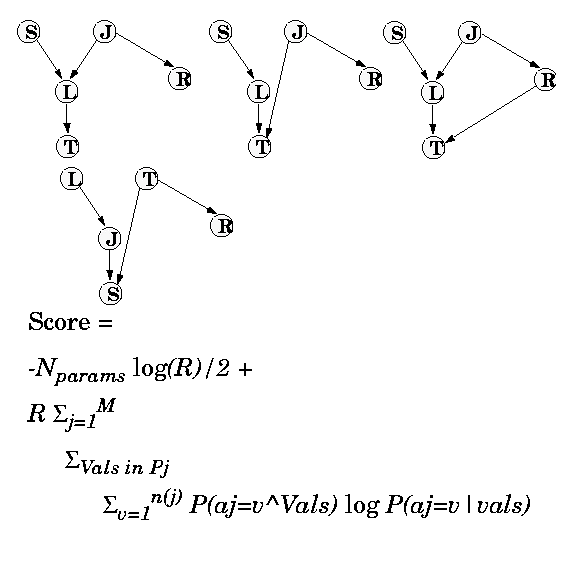
the net

Score 

Scoring a Numsbetrrouf ncont- ure redundant

parameters defining

#Attributes



 *P*(*Vk* )*P*( *X j*  *v* | *Vk* ) log *P*( *X j*  *v* |*Vk* )

*v* 1

All these values estimated from data

Bayes Net Structure: Slide 7

Copyright © 2001, Andrew W. Moore



*k* 1

 num combinations

*m*  of parent vaules  (arityof *X j* )

* *R*

*j* 1

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)

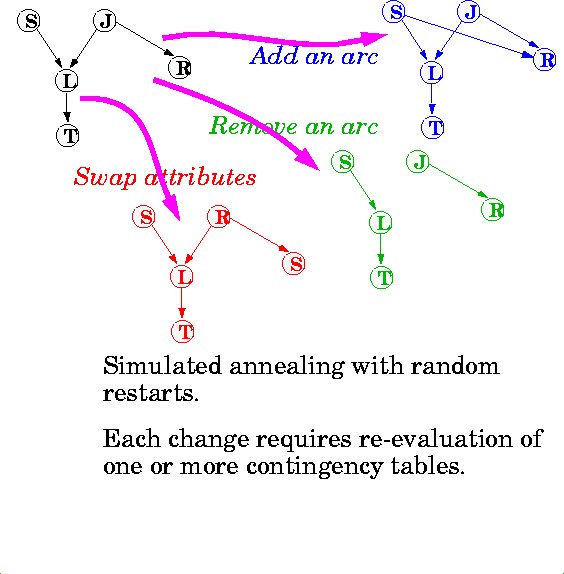
params log *R*

2

* *N*

Score 

Scoring a structure



Bayes Net Structure: Slide 8

Copyright © 2001, Andrew W. Moore

Searching for structure with best score

Learning Methods until today

Predict category

Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh

Prob- ability

Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE

Predict real no.

Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 9

Regressor

Classifier

Learning Methods added today

Predict category

Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh

Prob- ability

Predict real no.

Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve 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

Regressor

Classifier

Inputs Inputs Inputs

Inputs Inputs Inputs

|  |  |  |
| --- | --- | --- |
|  | Density Estimator |  |
|  |
|  |
|  |

|  |  |  |
| --- | --- | --- |
|  | Density Estimator |  |
|  |
|  |  |

But also, for free…

Predict category

Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh, Bayes Net Based BC

Prob- ability

Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE, Bayes Net Structure Learning

Predict real no.

Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 11

Regressor

Classifier

And a new operation…

Inference Joint DE, Bayes Net Structure Learning

Engine Learn P(E1|E2)

Predict category

Prob- ability

Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naïve BC, N.Neigh, Bayes Net Based BC

Joint DE, Naïve DE, Gauss/Joint DE, Gauss Naïve DE, Bayes Net Structure Learning

Predict real no.

Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore

Bayes Net Structure: Slide 12

Regressor

Classifier

Inputs Inputs Inputs Inputs

Inputs Inputs Inputs

|  |  |  |
| --- | --- | --- |
|  | Density Estimator |  |
|  |
|  |
|  |

|  |  |  |
| --- | --- | --- |
|  | Density Estimator |  |
|  |
|  |  |