

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 lecture slides: <http://www.cs.cmu.edu/~awm/lectures/>. 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

$P(S) = 0.3$
 $P(J) = 0.6$
 $P(L | J \wedge S) = 0.05$
 $P(L | J \wedge \sim S) = 0.1$
 $P(L | \sim J \wedge S) = 0.1$
 $P(L | \sim J \wedge \sim S) = 0.2$
 $P(R | J) = 0.3$
 $P(R | \sim J) = 0.6$
 $P(T | L) = 0.3$
 $P(T | \sim L) = 0.8$

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 2

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

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

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
:	:	:	:	:

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 3

Estimating Probability Tables

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

Counts: $ct(S)$, $ct(J)$, $ct(L, J, S)$, $ct(L, \sim S)$, $ct(R, J)$

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
:	:	:	:	:

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 4

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

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

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 5

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

All these values estimated from data

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

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 6

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)

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

All these values estimated from data

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 7

Searching for structure with best score

Simulated annealing with random restarts.

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

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 8

Learning Methods until today

Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh
Inputs → Density Estimator → Probability	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE
Inputs → Regressor → Predict real no.	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

Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh
Inputs → Density Estimator → Probability	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive 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

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 10

But also, for free...

Inputs → Classifier → Predict category	Dec Tree, Sigmoid Perceptron, Sigmoid N.Net, Gauss/Joint BC, Gauss Naive BC, N.Neigh, Bayes Net Based BC
Inputs → Density Estimator → Probability	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE, Bayes Net Structure Learning
Inputs → Regressor → Predict real no.	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...

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 Naive BC, N.Neigh, Bayes Net Based BC
Inputs → Density Estimator → Probability	Joint DE, Naive DE, Gauss/Joint DE, Gauss Naive DE, Bayes Net Structure Learning
Inputs → Regressor → Predict real no.	Linear Regression, Quadratic Regression, Perceptron, Neural Net, N.Neigh, Kernel, LWR

Copyright © 2001, Andrew W. Moore Bayes Net Structure: Slide 12