



TAN Trees

Suggested reading:

- Bayesian Network Classifiers, Friedman et al., see class website

Machine Learning 10-601

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

February 27, 2008

What is form of Bayes net for Naïve Bayes

How can we learn structure, in way that
assures polynomial time learning?
TAN tree (Tree Augmented Network)

TAN Trees

TAN Tree: a Bayes net similar to naïve Bayes for $P(Y|X_1, \dots, X_n)$, except:

- Impose a tree structure over the X_i nodes (at most one parent per X_i)

We can learn the best such tree in time polynomial in size of tree and data

TAN Tree Algorithm finds TAN tree that maximizes data likelihood, in poly time

$$I_P(\mathbf{X}; \mathbf{Y} | \mathbf{Z}) = \sum_{\mathbf{x}, \mathbf{y}, \mathbf{z}} P(\mathbf{x}, \mathbf{y}, \mathbf{z}) \log \frac{P(\mathbf{x}, \mathbf{y} | \mathbf{z})}{P(\mathbf{x} | \mathbf{z}) P(\mathbf{y} | \mathbf{z})}$$

The **Construct-TAN** procedure consists of five main steps:

1. Compute $I_{\hat{P}_D}(A_i; A_j | C)$ between each pair of attributes, $i \neq j$.
2. Build a complete undirected graph in which the vertices are the attributes A_1, \dots, A_n . Annotate the weight of an edge connecting A_i to A_j by $I_{\hat{P}_D}(A_i; A_j | C)$.
3. Build a maximum weighted spanning tree.
4. Transform the resulting undirected tree to a directed one by choosing a root variable and setting the direction of all edges to be outward from it.
5. Construct a TAN model by adding a vertex labeled by C and adding an arc from C to each A_i .

Experimental results: NB, Bayes Nets, TAN Trees

	Data set	NB	BN	TAN ^s
1	australian	86.23+-1.10	86.23+-1.76	84.20+-1.24
2	breast	97.36+-0.50	96.92+-0.63	96.92+-0.67
3	chess	87.15+-1.03	95.59+-0.63	92.31+-0.82
4	cleve	82.76+-1.27	81.39+-1.82	81.76+-0.33
5	corral	85.88+-3.25	97.60+-2.40	96.06+-2.51
6	crx	86.22+-1.14	85.60+-0.17	85.76+-1.16
7	diabetes	74.48+-0.89	75.39+-0.29	75.52+-1.11
8	flare	79.46+-1.11	82.74+-1.90	82.27+-1.86
9	german	74.70+-1.33	72.30+-1.57	73.10+-1.54
10	glass	69.66+-1.85	55.57+-5.39	67.78+-3.43
11	glass2	79.17+-1.71	75.49+-2.47	77.92+-1.11
12	heart	81.48+-3.26	82.22+-2.46	83.33+-2.48
13	hepatitis	91.25+-1.53	91.25+-4.68	91.25+-2.50
14	iris	93.33+-1.05	94.00+-1.25	94.00+-1.25
15	letter	74.96+-0.61	75.02+-0.61	85.86+-0.49
16	lymphography	79.72+-1.10	75.03+-1.58	85.03+-3.09
17	mofn-3-7-10	86.43+-1.07	85.94+-1.09	91.11+-0.89
18	pima	75.51+-1.63	75.00+-1.22	75.52+-1.27
19	satimage	81.75+-0.86	59.20+-1.10	87.20+-0.75
20	segment	91.17+-1.02	93.51+-0.89	95.58+-0.74
21	shuttle-small	98.34+-0.29	99.17+-0.21	99.53+-0.15
22	soybean-large	91.29+-0.98	58.54+-4.84	92.17+-1.02
23	vehicle	58.28+-1.79	61.00+-2.02	69.63+-2.11
24	vote	90.34+-0.86	94.94+-0.46	93.56+-0.28
25	waveform-21	77.89+-0.61	69.45+-0.67	78.38+-0.60