Master of Technology in Knowledge Engineering

Bayesian Modeling

Dr. Barry Shepherd
Institute of Systems Science
National University of Singapore
Email: barryshepherd@nus.edu.sg

"Do Doctors Understand Test Results?"

BBC NEWS MAGAZINE
6 July 2014

In 2006 & 2007 Gigerenzer (a prominent statistician) gave a series of workshops to more than 1,000 practising gynaecologists, and kicked off every session with the same question:

Question
"A 50-year-old woman, no symptoms, has routine mammography screening.
She tests positive, is alarmed, and wants to know if she has breast cancer for certain or what the chances are.
You know nothing else about this woman.
How many women who test positive actually have breast cancer? "

Background knowledge:
- *The probability a woman has breast cancer is 1% ("prevalence")*
- *If a woman has breast cancer, the probability of a positive test is 90% ("sensitivity")*
- *If a woman doesn't have breast cancer the probability of a positive test is 9% ("false alarm rate")*

What is the answer?
- *nine in 10*
- *eight in 10*
- *one in 10*
- *one in 100*

Only 21% got the answer correct!

Bayes Theorem

- P(event | observation) = $\frac{\text{P(observation | event) * P(event)}}{\text{P(observation)}}$

More familiarly..

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

- For breast cancer
  - $P(\text{cancer} \mid \text{positive test}) = \underbrace{P(\text{positive test} \mid \text{cancer})}_{0.9\ \textit{(known)}} * \underbrace{P(\text{cancer})}_{0.01\ \textit{(known)}} / \underbrace{P(\text{positive test})}_{?}$
- How do we get P(positive test)?
  - P(observation) = Σ_{events} P(observation | event) * P(event) … *using law of total probability*
  - P(positive test) = P(positive |cancer) *P(cancer) + P(positive | not cancer)P(not cancer)
    = 0.9*0.01 + 0.09*(1-0.01) = 0.098

if $\{B_n : n = 1, 2, 3, \ldots\}$ is a finite partition of a sample space

$$\Pr(A) = \sum_n \Pr(A \cap B_n)$$

$$\Pr(A) = \sum_n \Pr(A \mid B_n)\Pr(B_n)$$

Law of total probability

- Hence
  - P(cancer | positive test) = 0.9 * 0.01 / 0.098 = 0.0918 ~ about 1 in 10

Yellow Cab Problem

- In a town there are 85% yellow cabs and 15% white ones
- A pedestrian is knocked down by a cab that ran away
- A witness says the cab was white
- Witness undergoes sight color tests and gives correct color only 4 times in 5
- Is the guilty cab white or yellow?

Yellow Cab Problem

- P(cab was white | witness saw white)

  = P(saw white | was white) P(was white) / P (saw white)

  = 0.8 * 0.15 / (0.8 * 0.15 + 0.2 * 0.85)

  = 0.12 / (0.12 + 0.17) = 0.4137

  Where:

  P(saw white) = P(saw white | was white) * P(white) + P(saw white| not white)P(not white)

A More Complex Scenario

- A population has 50% boys and 50% girls
- There are two schools; Romeo and Juliette
- 75% of boy go to Romeo School, 75% of girls go to Juliette
- In each school, boys have a higher exam pass rate
  - In Romeo they had 40% pass rate compared to the girls 30% pass rate
  - In Juliet they had 80% pass rate compared to the girls 70% pass rate
- Who are clever overall? boys or girls?

The Monty Hall Problem

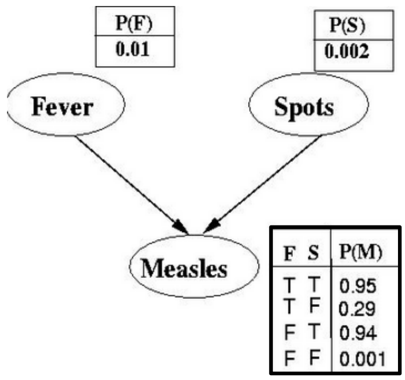
A gift is secretly placed under one of 3 hats – you must guess which one

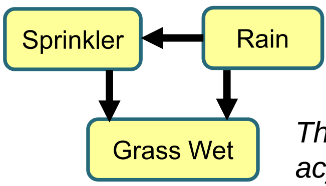
Rules….

- You choose one hat
- The organiser raises one of the remaining two
  (of course not the one with the gift if it is not under your hat)
- There are now two hats left, should you keep your initial choice?

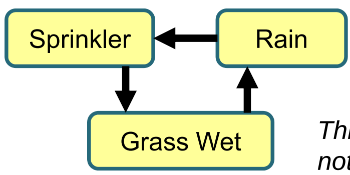
Bayesian Belief Networks

- Directed Acyclic Graphs (DAGs)…
- Represent a set of random variables and their conditional dependencies
- Nodes can be observable quantities, latent variables (not observable, inferred only), unknown parameters or hypotheses.
- Edges represent conditional dependencies; nodes that are not connected represent variables that are conditionally independent of each other





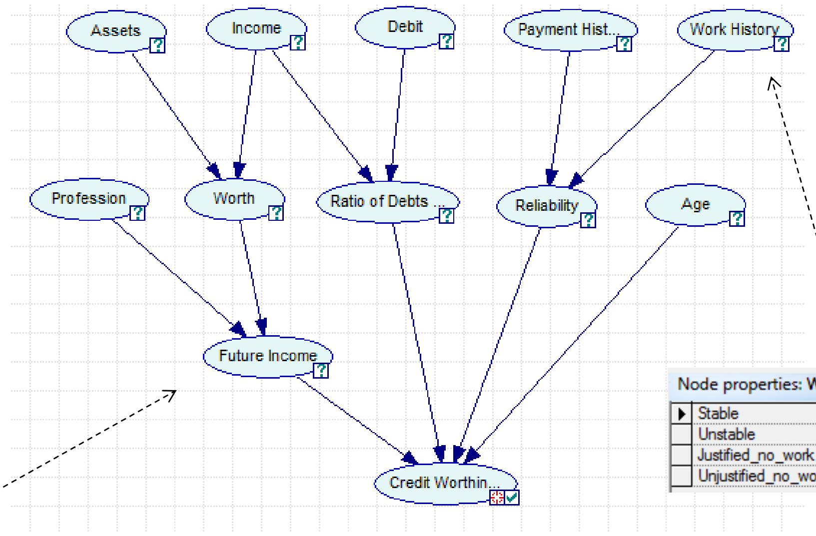
This is acyclic



This is not!

Example: Credit Worthiness

A simple network for assessing credit worthiness of an individual, developed by Gerardina Hernandez as a class homework at the University of Pittsburgh.
Note that all parentless nodes are described by uniform distributions. This is a weakness of the model, although it is offset by the fact that all these nodes will usually be observed and the network will compute the probability distribution over credit worthiness correctly. Another element of this model is that only the node CreditWorthiness is of interest to the user and is designated as a target.



Node properties: Work History

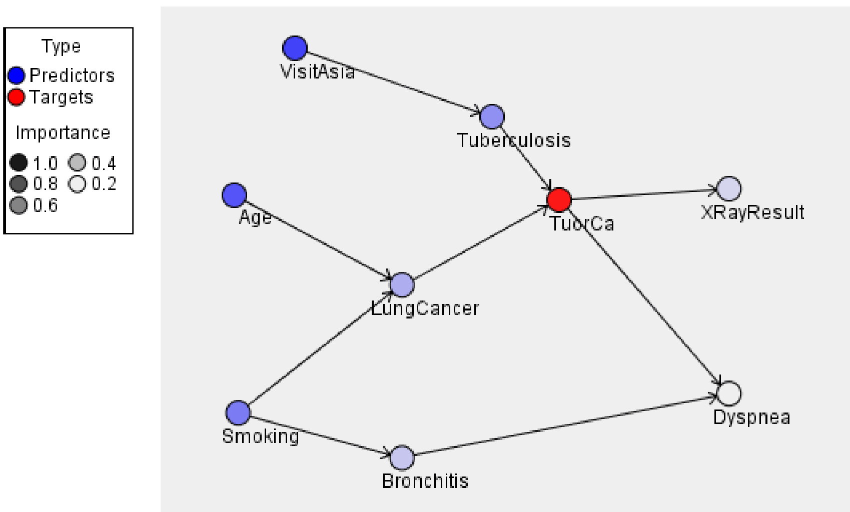
Stable	0.25
Unstable	0.25
Justified_no_work	0.25
Unjustified_no_work	0.25

Node properties: Future Income

Worth	High			Medium			Low		
Profession	High_inco...	Medium_in...	Low_incom...	High_inco...	Medium_in...	Low_incom...	High_inco...	Medium_in...	Low_incom...
Promising	0.99	0.8	0.6	0.85	0.6	0.4	0.8	0.4	0.01
Not_promissing	0.01	0.2	0.4	0.15	0.4	0.6	0.2	0.6	0.99

GeNIe & SMILE
University of Pittsburgh

Example: Asia Travel



From Lauritzen and Spiegelhalter (1988)

Uses:

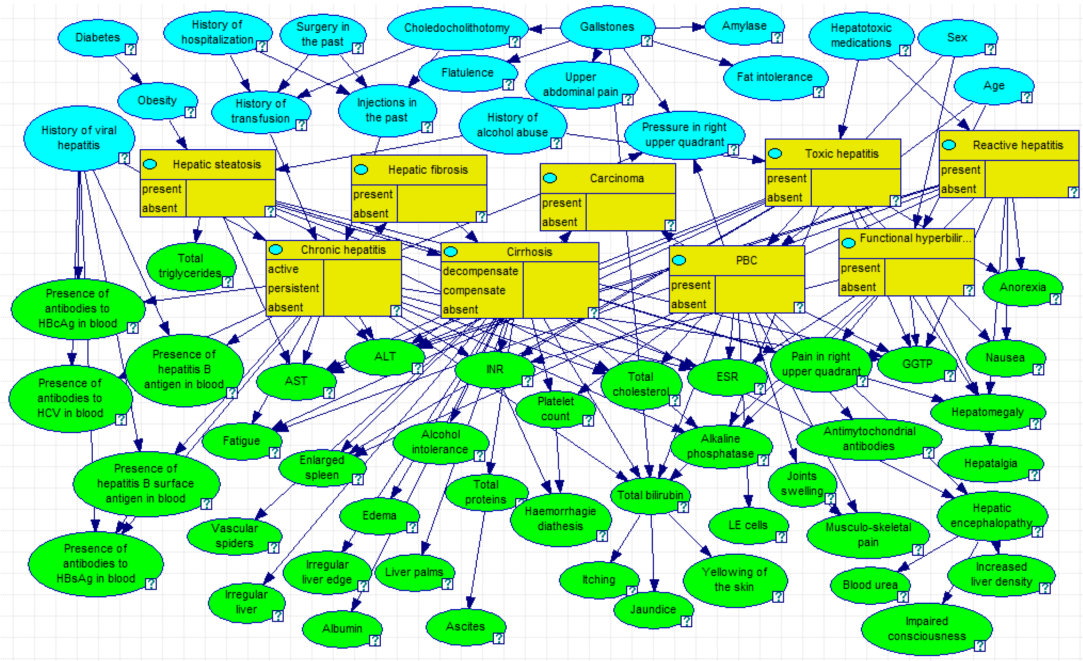
- **Prediction:** Given observations and evidence what is the target likelihood?
- **Diagnosis:** Given symptoms what are the probable causes?

See https://www.norsys.com/tutorials/netica/secA/tut_A1.htm

Having a target is not necessary….

Exploration: alter observations and evidence levels to explore the interactions between the domain entities

Example: Liver Disease Diagnosis



Hepar II is a Bayesian network model for the diagnosis of liver disorders, generously contributed to the community by Agnieszka Onisko.
The primary reference for the Hepar II model is her doctoral dissertation (available at: http://aragorn.pb.bialystok.pl/~aonisko): Agnieszka Onisko. Probabilistic Causal Models in Medicine: Application to Diagnosis of Liver Disorders. Ph.D. Dissertation, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Science, Warsaw, March 2003.

The orange colored nodes represent diseases, the blue nodes represent history, including risk factors, and green nodes represent observations and test results.
We suggest that you explore the model using the Test View of the diagnostic interface (a menu choice in the Diagnosis Menu, a button with green horizontal bars or simply F7 function key).

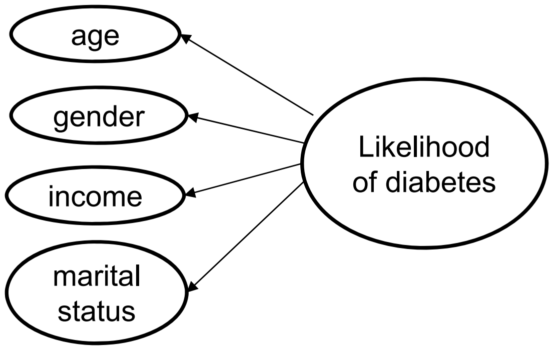
How do we build Bayesian Networks?

- Manual Build
  - Most dedicated Bayesian Net tools allow the user to paste nodes onto the screen and connect them according to their own domain knowledge
  - They users must them input their own estimates of the a-priori & conditional probability distributions
  - Manual build can be difficult for large networks
  - BayesiaLab has tools that allow a group of experts to brainstorm a network
- Machine Learning Algorithms
  - Many algorithms: e.g. SPSS modeler has “Tree Augmented Naïve Bayes “ (TAN) and “Markov Blanket” learning algorithms
  - Learned networks also allow for knowledge discovery ~ visualisation and analysis of the learned graph structure can yield insights into the problem domain.

How to Learn Bayesian Networks?

- Easier ~ Manually define the structure and learn the probabilities from data
- Harder ~ learn the structure as well as the probabilities from data

- Simplest form of Bayes Net to learn = Naïve Bayes Classifier



The target node is the parent of all other nodes

Very robust predictors – even if the conditional independence requirement does not hold

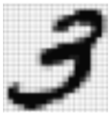
Naïve Bayes Classifier

- The Naïve Bayes Classifier
  - Given a set of observations (a_1, a_2, a_3, ..) pick between a set of classes C
  - Calculate P(Class=c_j | Observations) for each class j and pick the class with the biggest value
- From Bayes Theorem we know that
  - P(C|A) = P(A|C) * P(C)/ P(A)
  - Hence we pick the class j that maximises P(A | C=c_j) * P(C=c_j)
- We use the training data to estimate the prior and conditional probabilities:
  - $P(C=c_j) = \dfrac{\#\text{examples of class } c_j}{\#\text{total examples in training set}}$
  - $P(A \mid C=c_j) = P(A_1=a_1 \ \& \ A_2=a_2 \ldots \mid C=c_j) = P(A_1=a_1 \mid C=c_j) * P(A_2=a_2 \mid C=c_j) * \ldots$

    (this assumes that the attributes A_i are conditionally independent)
  - $P(A_i=a_i \mid C=c_j) = \dfrac{\#\text{examples with } A_i = a_i \text{ and class } C = c_j}{\#\text{examples of class } c_j}$

Why require Conditional Independence?

- $P(A_1=a_1 \;\&\; A_2=a_2 \ldots | C=c_j)$ is often hard to estimate by frequency counting due to the huge number of combinations of the attributes (observations) A_i
- E.g. Image recognition of handwritten numerical digits (0->9)



Assume 20*20 pixel images (black and white only) => 2^{400} possible images

E.g. For the above image we would need:

$$P(Pixel_1 = 0 \;\&\; Pixel_2=0 \;\&\; \ldots\; Pixel_{17} = 1 \;\&\; \ldots\; Pixel_{400}=0 \,|\, C=3)$$

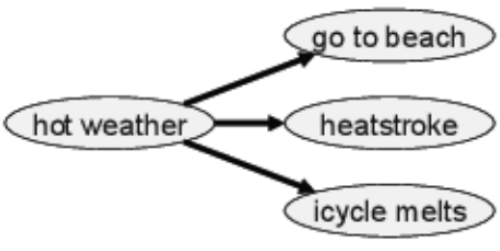
- Conditional Independence allows us to greatly reduce the number of probabilities we need to estimate

$$\underbrace{P(A_1=a_1 \;\&\; A_2=a_2 \ldots | C=c_j)}_{2^{400}\text{ probabilities needed to cover all cases}} = \underbrace{P(A_1=a_1 | C=c_j) * P(A_2=a_2 | C=c_j) * \ldots}_{400\text{ probabilities needed}}$$

- E.g.

Is Conditional Independence Valid?

- Consider an example….
- Going to the beach and heatstroke are not independent, they are both more likely in times of hot weather
  - P(Beach, Heatstroke) > P(Beach) * P(Heatstroke)
- But they may be independent if we know that the weather is hot
  - E.g. if we know its hot then knowledge about going to the beach does not impact probability of heatstroke
  - P(Beach, Heatstroke) = P(Beach | Hot weather) * P(Heatstroke | Hot Weather)



Example © Amos Storkey, 2007

Naïve Bayes Classification: An Example

- Given a set of people with known symptoms and known condition (flu or not)

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

- Does the following new patient have the flu?

Chills	Runny Nose	Headache	Fever	Flu?
Y	N	Mild	N	?

** Scenario from http://www.youtube.com/watch?v=ZAfarappAO0*

Naïve Bayes Classification: An Example

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

P(Flu=Y)	0.625	P(Flu=N)	0.375
P(chills=Y|flu=Y)	0.6	P(chills=Y|flu=N)	0.333
P(chills=N|flu=Y)	0.4	P(chills=N|flu=N)	0.666
P(runny nose=Y|flu=Y)	0.8	P(runny nose=Y|flu=N)	0.333
P(runny nose=N|flu=Y)	0.2	P(runny nose=N|flu=N)	0.666
P(headache=Mild|flu=Y)	0.4	P(headache=Mild|flu=N)	0.333
P(headache=No|flu=Y)	0.2	P(headache=No|flu=N)	0.3333
P(headache=Strong|flu=Y)	0.4	P(headache=Strong|flu=N)	0.333
P(fever=Y|flu=Y)	0.8	P(fever=Y|flu=N)	0.333
P(fever=N|flu=Y)	0.2	P(fever=N|flu=N)	0.666

Chills	Runny Nose	Headache	Fever	Flu?
Y	N	Mild	N	?

$$\underset{c}{\operatorname{argmax}}\ p(C=c)\prod_{i=1}^{n} p(F_i=f_i|C=c).$$

P(flu=Y) * P(chills=Y | flu=Y) * P(runny nose =N | flu=Y) * P(headache=mild | flu=Y) * P(fever=N | flu=Y) = 0.006

vs

P(flu=N) * P(chills=Y | flu=N) * P(runny nose =N | flu=N) * P(headache=mild | flu=N) * P(fever=N | flu=N) = 0.0185

Naïve Bayes: Another Example

- Given some observations ....

sex	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

....is this person male or female?

sex	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$\underset{c}{\operatorname{argmax}}\ p(C=c)\prod_{i=1}^{n} p(F_i=f_i|C=c).$$

How do we handle numerical observations?

- For small training sets we use Gaussian assumption (assume inputs are normally distributed)
- For larger training sets we must use binning (discretization)

Naïve Bayes: Handling Numerical Inputs

- To use the Gaussian assumption we first compute the mean and variance

sex	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033e-02	176.25	1.2292e+02	11.25	9.1667e-01
female	5.4175	9.7225e-02	132.5	5.5833e+02	7.5	1.6667e+00

- Then apply the Gaussian formula to the observation to get the conditionals:

height (feet)	weight (lbs)	foot size(inches)
6	130	8

$$P(male) = 0.5$$

$$\text{P(height = 6 | male)} = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789,$$

- Finally

$P(male) = 0.5$ $p(\text{height}\mid\text{male}) =$ $p(\text{weight}\mid\text{male}) = 5.9881\cdot 10^{-6}$ $p(\text{foot size}\mid\text{male}) = 1.3112\cdot 10^{-3}$ $\text{posterior numerator (male)} = \text{their product} = 6.1984\cdot 10^{-9}$	$P(\text{female}) = 0.5$ $p(\text{height}\mid\text{female}) = 2.2346\cdot 10^{-1}$ $p(\text{weight}\mid\text{female}) = 1.6789\cdot 10^{-2}$ $p(\text{foot size}\mid\text{female}) = 2.8669\cdot 10^{-1}$ $\text{posterior numerator (female)} = \text{their product} = 5.3778\cdot 10^{-4}$

Naïve Bayes: Handling Zero Counts

- Observations that occur very infrequently can cause problems….

- Separate spam from valid email, attributes = words

D1: "send us your password" spam
D2: "send us your review" ham
D3: "review your password" ham
D4: "review us" spam
D5: "send your password" spam
D6: "send us your account" spam

P (spam) = 4/6 P (ham) = 2/6

spam	ham	
2/4	1/2	password
1/4	2/2	review
3/4	1/2	send
3/4	1/2	us
3/4	1/2	your
1/4	0/2	account

new email: "review us now"

$$P(\text{review us}\,|\text{spam})=P(0,1,0,1,0,0\,|\text{spam})=(1-\frac{2}{4})(\frac{1}{4})(1-\frac{3}{4})(\frac{3}{4})(1-\frac{3}{4})(1-\frac{1}{4})$$

$$P(\text{review us}\,|\text{ham})=P(0,1,0,1,0,0\,|\text{ham})=(1-\frac{1}{2})(\frac{2}{2})(1-\frac{1}{2})(\frac{1}{2})(1-\frac{1}{2})(1-\frac{0}{2})$$

$$P(\text{ham}\,|\text{review us})=\frac{0.0625\times 2/6}{0.0625\times 2/6+0.0044\times 4/6}=0.87$$

See http://www.inf.ed.ac.uk/teaching/courses/iaml/slides/naive-2x2.pdf

Naïve Bayes: The Zero-Count Problem

- It is possible that a particular attribute value in the test/evaluation set never occurs with a class in the training set. This will result in a zero probability for that attribute class. E.g.
  - P("account" | ham) = 0/2, hence any email containing "account" must be spam!
- In general
  - if P(Observation | Class) = 0 then P(Class | Observation) = 0 *but this is often not true!*
- A small-sample correction is often used to prevent zero probabilities:

$$P(A_i \mid C_j) = (n_{ij} + \lambda) / (n_j + \lambda\, m_i)$$

where λ is commonly set to be 1/n and m_i is the number of values of attribute j

For P(account | ham) we set λ =2 (the number of instances of class = ham)

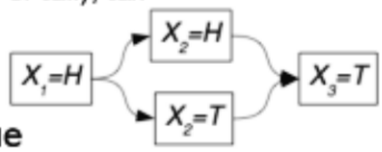
$$P(w|c)=\frac{num(w,c)+\epsilon}{num(c)+2\epsilon}$$

- In spam detection this is a common problem
  - Zipf's law –approx. 50% of words occur only once

Naïve Bayes: Handling Missing Values

- Bayesian Nets are very robust with respect to missing values
- We can ignore missing values when computing the probability estimates
  - $P(A_1=a_1 \ldots\ \mathbf{A_j=?}\ \ldots\ A_n=a_b \mid C=c_j) = P(A_1=a_1 \mid C=c_j) * \ldots * 1 * \ldots P(A_n=a_n \mid C=c_j)$

- Ex: three coin tosses: Event = $\{X_1=H, X_2=?, X_3=T\}$
  - event = head, unknown (either head or tail), tail
  - event = {H,H,T} + {H,T,T}
  - *P*(event) = P(H,H,T) + P(H,T,T)



- General case: X_j has missing value

$$P(x_1 \ldots \boxed{x_j} \ldots x_d \mid y) = P(x_1 \mid y) \cdots \boxed{P(x_j \mid y)} \cdots P(x_d \mid y)$$

$$\sum_{\boxed{x_j}} P(x_1 \ldots \boxed{x_j} \ldots x_d \mid y) = \sum_{x_j} P(x_1 \mid y) \cdots \boxed{P(x_j \mid y)} \cdots P(x_d \mid y)$$

$$= P(x_1 \mid y) \cdots \boxed{\sum_{x_j} P(x_j \mid y)} \cdots P(x_d \mid y)$$

$$= P(x_1 \mid y) \cdots \boxed{1} \cdots P(x_d \mid y)$$

See http://www.inf.ed.ac.uk/teaching/courses/iaml/slides/naive-2x2.pdf

Improvements on Naïve Bayes Learning

- Augmented Naïve Bayes
- Tree Augmented Naïve Bayes
- Markov Blanket
- Augmented Markov Blanket
- Semi-Structured Learning
- Expert Guided Learning
- Others…

An Example

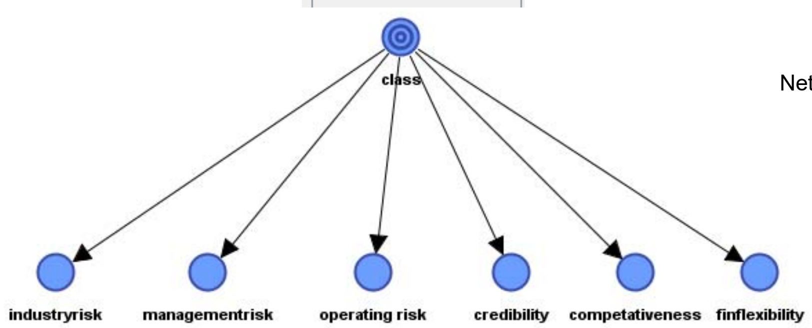
- Predicting Corporate Bankruptcy *
- 6 discrete predictors (values = positive, average, negative)
- one target (class) (values = B/NB)
- 250 examples:
  - 107 bankrupt (B)
  - 143 non-bankrupt (NB)

MAIN PARAMETERS	SUB PARAMETERS
Industry Risk(IR) (Myoung and Ingoo,2003)	▪ Governmental Policies and International Agreements (GPI) ▪ Cyclicality (CY) ▪ Degree of Competition (DEG) ▪ Price and Stability of the Market Supply(PSM) ▪ Size and Growth of Market Demand (SGM) ▪ Sensitivity to changes in Macroeconomics (SEN) ▪ Domestic and International competitive Power (DOM-POW) ▪ Product Life Cycle (PLC)
Management Risk(MR) (Myoung and Ingoo,2003)	▪ Ability and Competence of management (ACM) ▪ Stability of Management (SM) ▪ Relationship between Management/Owner (REL-MO) ▪ Human Resource Management (HRM) ▪ Growth Process/Business Performance (GP/BP) ▪ Short and long term Business Planning, Achievement and Feasibility (SL-TERM AF)
Financial Flexibility (FF) (Myoung and Ingoo,2003)	▪ Direct Financing (DF) ▪ Indirect Financing (IF) ▪ Other financing [Affiliates, Owner, Third Parties] (OF)
Credibility(CR) (Myoung and Ingoo,2003)	▪ Credit History (CH) ▪ Reliability of Information (RI)
Competitiveness (CO) (Myoung and Ingoo,2003)	▪ Market Position (MP) ▪ Differentiated Strategy (DIFF-S)
Operating Risk (OR) (Myoung and Ingoo,2003)	▪ Stability and Diversity of Procurement ▪ Stability of Transaction ▪ Performance of Production ▪ Prospectus for Demand for Product and Service ▪ Sales Diversification ▪ Sales Price and Settlement Condition ▪ Effectiveness of Sales Network

** See "The discovery of experts decision rules from qualitative bankruptcy data using genetic algorithms" by Myoung-Jong Kim*, Ingoo Han.*

Naïve Bayes

B	NB
42.800	57.200



Network learned using BayesiaLab

P(Management risk | class)

class	A	N	P
B	21.495	68.224	10.280
NB	32.168	32.168	35.664

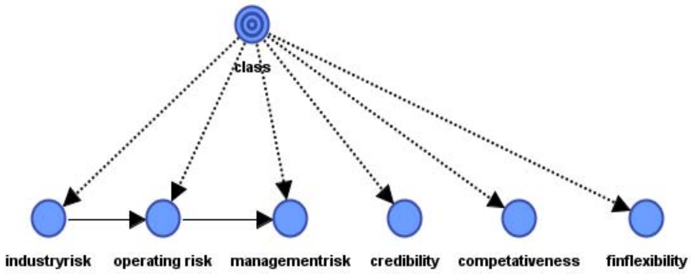
Node Selection: managementrisk

class	A	N	P
B	23	73	11
NB	46	46	51

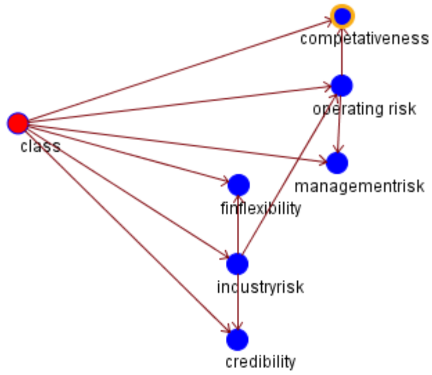
Occurrences managementrisk

Augmented Naïve Bayes

- Allows connections between child nodes, hence relaxes the conditional independence constraint (Tree Augmented Naïve Bayes is similar but faster)



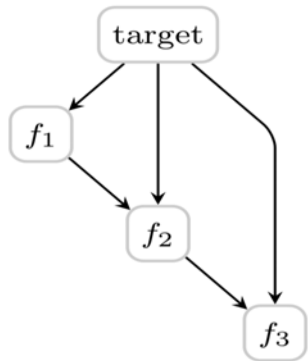
Augmented & Tree-Augmented Naïve Bayes - From BayesiaLab



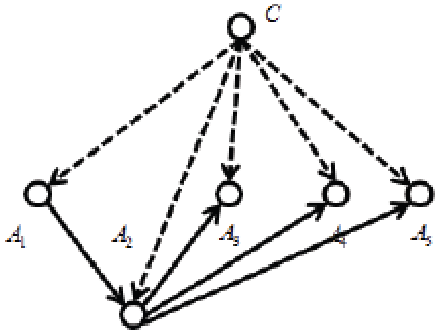
Tree-Augmented Naïve Bayes - From SPSS Modeler

Tree Augmented Naïve Bayes (TAN)

- The class variable has no parent and each other attribute has as parents the class variable and at most one other attribute
- TAN generally outperforms Naïve Bayes in terms of accuracy

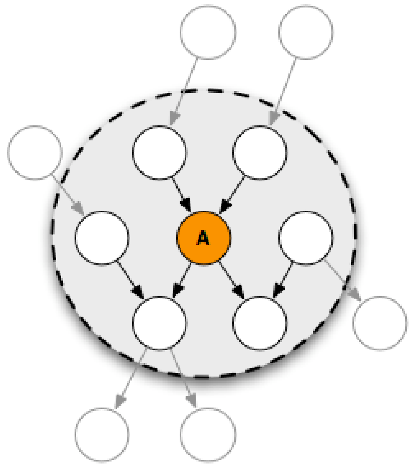


- During learning **one** of the attributes (input variables) is selected as the root variable, then all other attributes are connect to it in a directed tree (the direction is from the root to the leaves).
  - The structure of the tree is found by maximising the conditional mutual information between each pair of attributes (finding the maximal spanning tree)



Markov Blankets

- Markov Blanket ~ selects the nodes in the dataset that contain the target variable's parents, its children, and its children's parents
- The Markov blanket of a node contains all the variables that shield the node from the rest of the network. This means that the Markov blanket of a node is the only knowledge needed to predict the behavior of that node.
- Markov Blankets help in studying how an attribute x "behaves" under the effect of other attributes in the domain, by providing 'shielding' information.

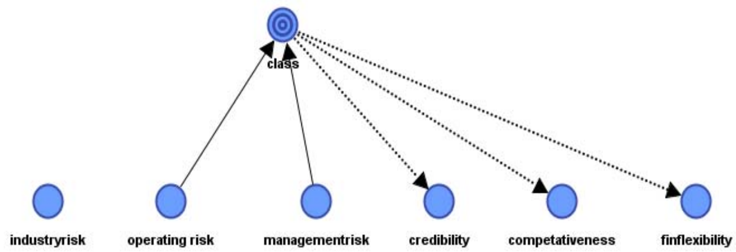


Markov Blanket learning is a supervised algorithm used to find a Bayesian Network that characterizes the Target node. *Regression*, *Decision Trees*, *Support Vector Machines*, *Neural Networks* are alternative supervised approaches, but they are all *Discriminative* models whereas the Markov Blanket algorithm returns a *Generative* model (this explains the redundancy of some features, but also allows to have more robust models, specifically when some values are missing)

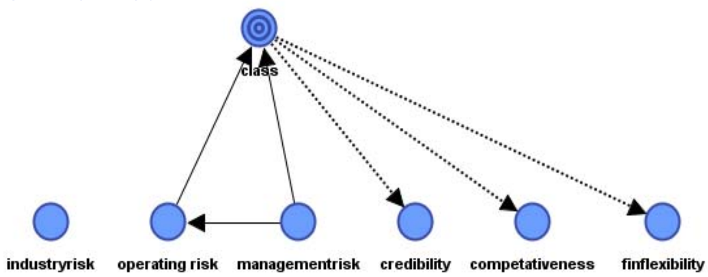
From: http://library.bayesia.com/display/FAQ/Markov+Blankets

Markov Blankets

- The Bankruptcy example as a Markov Blanket (BayesiaLab)....

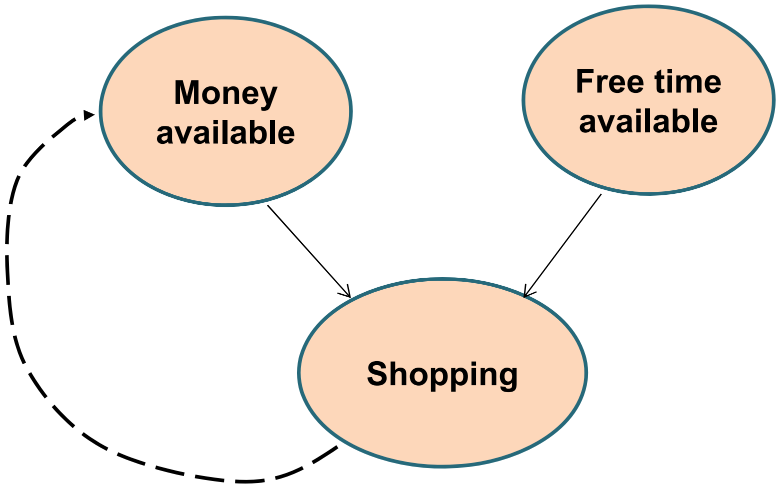


- As an augmented Markov Blanket..



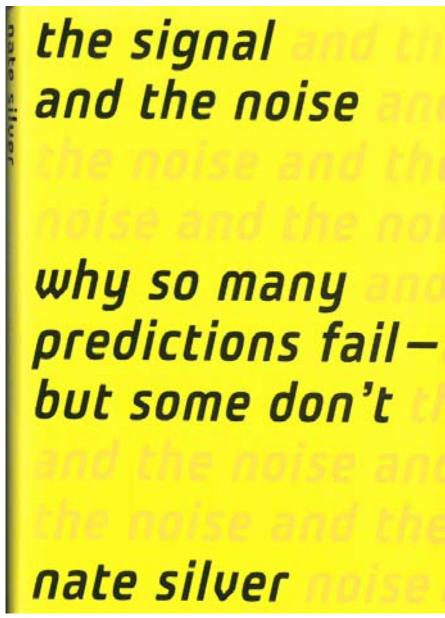
Dynamic Bayesian Networks

- How can we model this and still avoid a cyclic graph?
- How to handle the time dimension?



Refining a Prediction As Evidence Arrives

- Example: Prior to Sep 11th 2001, the probability of an attack on tall buildings in New York could be estimated (for example) as 0.005% (1 in 20,000)
- When the first plane crashed into WTC how much should that probability change?



Refining a Prediction As Evidence Arrives

How do the probabilities change when the first plane hits?

$$P(\text{attack} \mid \text{plane hits}) = \underbrace{P(\text{plane hits} \mid \text{attack})}_{y} * \underbrace{P(\text{attack})}_{x} / \underbrace{P(\text{plane hits})}$$

$$\underbrace{P(\text{plane hits} \mid \text{attack})}_{y} * \underbrace{P(\text{attack})}_{x} + \underbrace{P(\text{plane hits} \mid \text{no attack})}_{z} * \underbrace{P(\text{no attack})}_{(1-x)}$$

PRIOR PROBABILITY		
Initial estimate of how likely it is that terrorists would crash planes into Manhattan skyscrapers.	x	0.005%
A NEW EVENT OCCURS: FIRST PLANE HITS WORLD TRADE CENTER		
Probability of plane hitting if terrorists are attacking Manhattan skyscrapers.	y	100%
Probability of plane hitting if terrorists are *not* attacking Manhattan skyscrapers (i.e. an accident).	z	0.008%

there were 2 previous plane accidents in the previous 25,000 days (in 1945 & 1946) => 2 in 25,000 ~ 0.008%

Refining a Prediction As Evidence Arrives

Putting in the numbers to get the updated probability….

PRIOR PROBABILITY		
Initial estimate of how likely it is that terrorists would crash planes into Manhattan skyscrapers.	x	0.005%
A NEW EVENT OCCURS: FIRST PLANE HITS WORLD TRADE CENTER		
Probability of plane hitting if terrorists are attacking Manhattan skyscrapers.	y	100%
Probability of plane hitting if terrorists are *not* attacking Manhattan skyscrapers (i.e. an accident).	z	0.008%
POSTERIOR PROBABILITY		
Revised estimate of probability of terror attack, given first plane hitting World Trade Center.	$\frac{xy}{xy+z(1-x)}$	38%

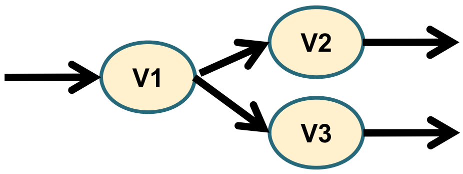
Refining a Prediction As Evidence Arrives

How do the probabilities change when the second plane hits?

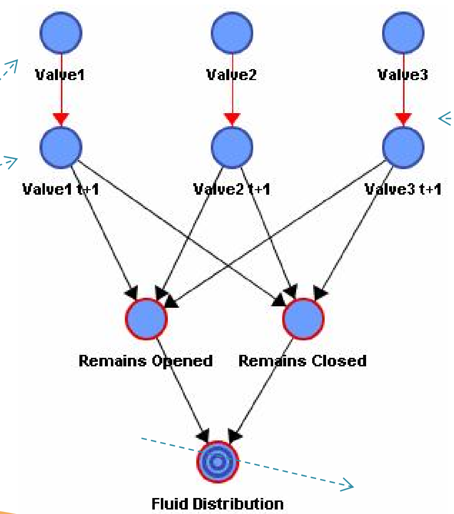
PRIOR PROBABILITY		
Revised estimate of probability of terror attack, given first plane hitting World Trade Center.	x	38%
A NEW EVENT OCCURS: SECOND PLANE HITS WORLD TRADE CENTER		
Probability of plane hitting if terrorists are attacking Manhattan skyscrapers.	y	100%
Probability of plane hitting if terrorists are *not* attacking Manhattan skyscrapers (i.e. an accident).	z	0.008%
POSTERIOR PROBABILITY		
Revised estimate of probability of terror attack, given second plane hitting World Trade Center.	$\frac{xy}{xy + z(1-x)}$	99.99%

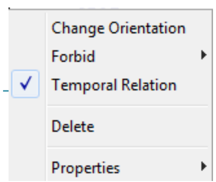
BayesiaLab – Dynamic Example

- A hydraulic system controls the flow of fluid by 3 valves. The valves open and close to maintain a constant fluid flow
- Each valve has a small probability of becoming faulty: remaining open (RO) or remaining closed (RC)
- How can we model and simulate the behavior of the system over time?



Model the transition over time by modeling the valves nodes at time t and t+1





Define some of the arcs as temporal

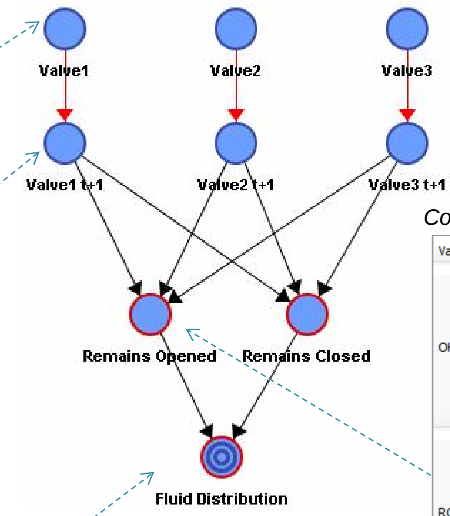
BayesiaLab – Dynamic Example

The valves are working properly at the start

OK	RO	RC
100.000	0.000	0.000

At time t+1 they have a small probability of failure. Once failed they remain in a fail state

Valve1	OK	RO	RC
OK	99.700	0.200	0.100
RO	0.000	100.000	0.000
RC	0.000	0.000	100.000



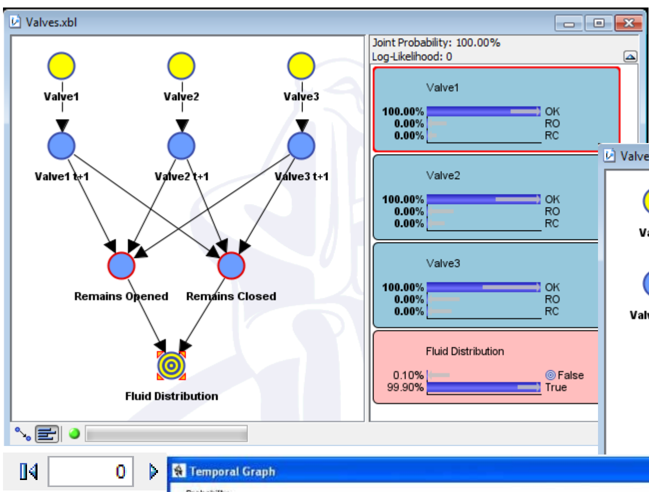
Remains Opened	Remains Closed	False	True
False	False	0.000	100.000
	True	100.000	0.000
True	False	100.000	0.000
	True	100.000	0.000

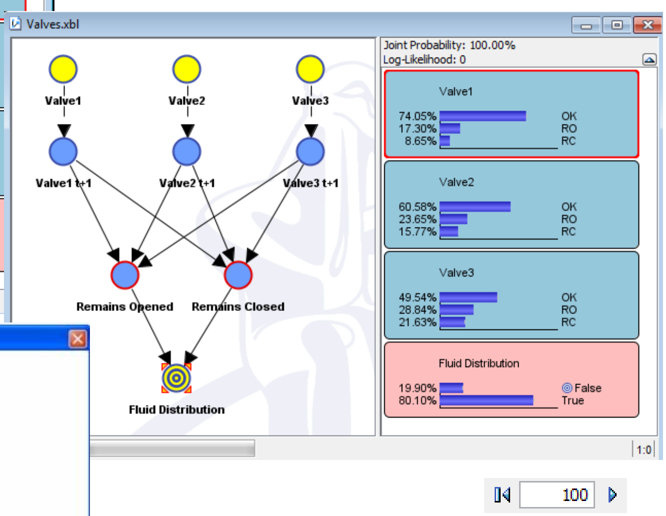
Conditional probabilities for "Remains Opened"

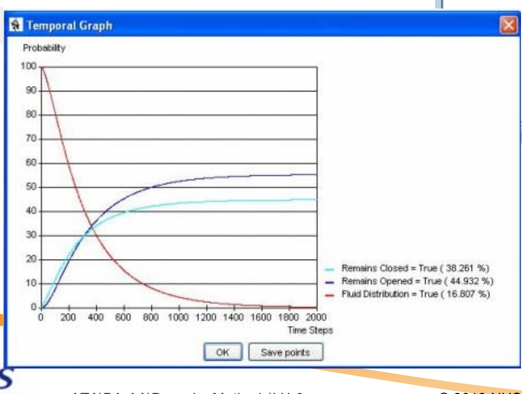
Valve1 t+1	Valve2 t+1	Valve3 t+1	False	True
OK	OK	OK	100.000	0.000
		RO	100.000	0.000
		RC	100.000	0.000
	RO	OK	100.000	0.000
		RO	100.000	0.000
		RC	100.000	0.000
	RC	OK	100.000	0.000
		RO	100.000	0.000
		RC	100.000	0.000
RO	OK	OK	100.000	0.000
		RO	0.000	100.000
		RC	100.000	0.000
	RO	OK	0.000	100.000
		RO	0.000	100.000
		RC	0.000	100.000
	RC	OK	100.000	0.000
		RO	0.000	100.000
		RC	100.000	0.000
RC	OK	OK	100.000	0.000
		RO	100.000	0.000
		RC	100.000	0.000
	RO	OK	100.000	0.000
		RO	100.000	0.000
		RC	100.000	0.000
	RC	OK	100.000	0.000
		RO	100.000	0.000
		RC	100.000	0.000

BayesiaLab – Dynamic Example

- In Evaluation mode you can step through time and see the behavior changes over time







www.bayesia.com/en/products/bayesialab/tutorial/chapter-5.php

Bayesian Workshop (A)

- Model the School Exam problem using GeNIe or BayesiaLab to determine whether boys or girls are clever on average
- Model the Monty Hall problem using GeNIe or BayesiaLab to determine which hat to choose

- Obtain GeNIe from: https://www.bayesfusion.com/

- Obtain BayesiaLab (30 day trial) from: http://www.bayesia.com/