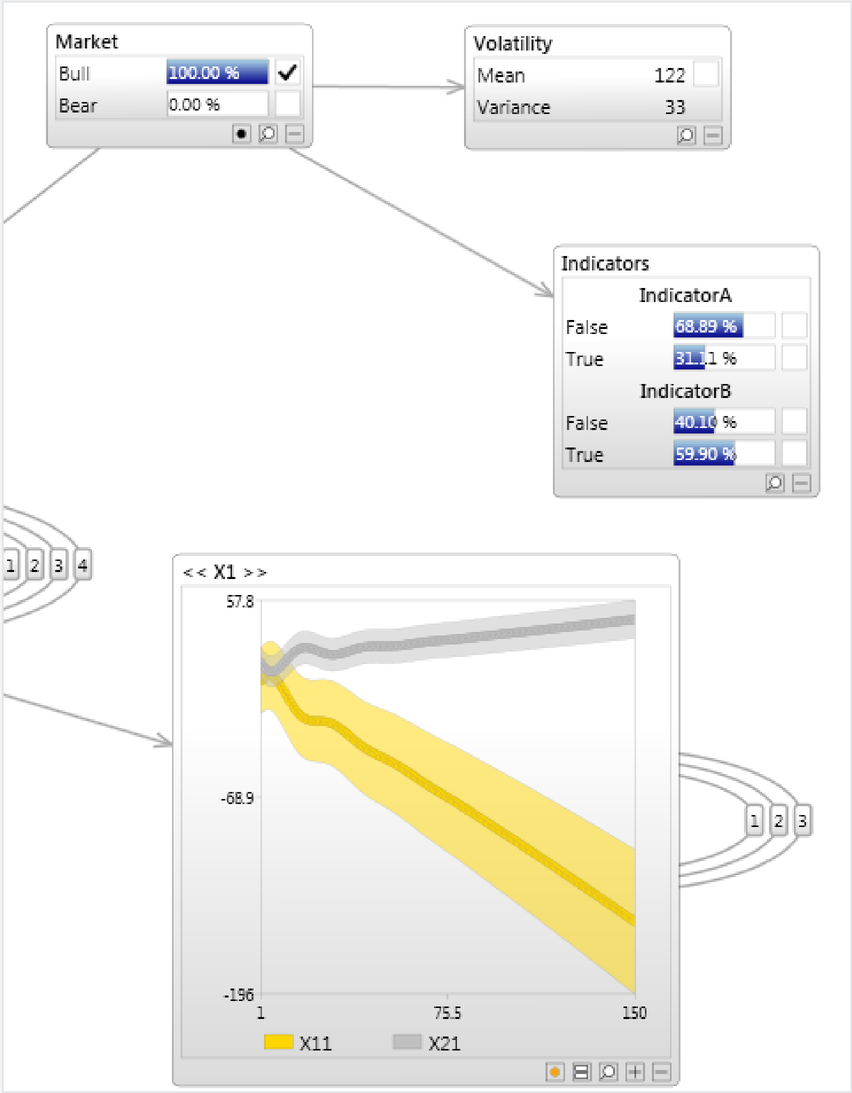
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer



Bayesian networks

Classification, segmentation, time series prediction and more.

John Sandiford

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Contents

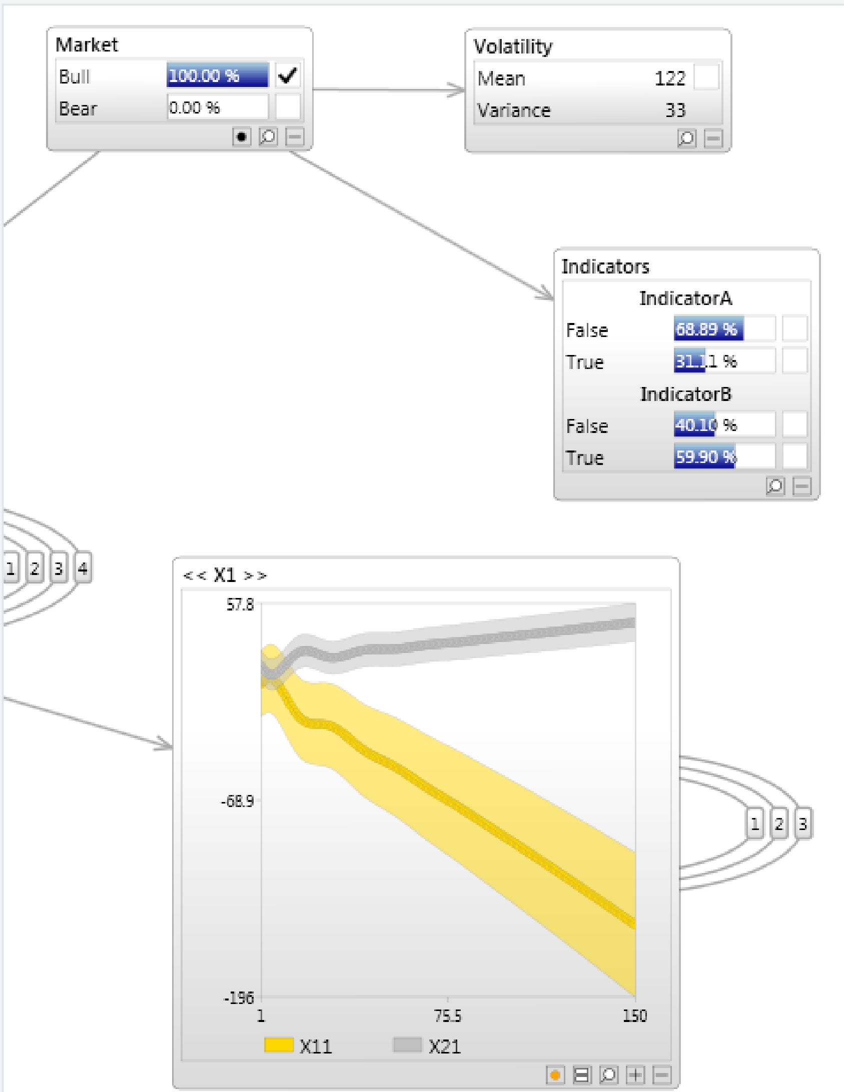
- Background
- Bayesian networks
- Classification / regression
- Clustering / segmentation
- Time series prediction

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Background

- Mathematics
- Algorithms
- Data Mining
- Machine Learning
- Artificial Intelligence
- Bayesian networks
  - Research (Imperial College)
  - Software

- BAE Systems
  - Future concepts
  - Ground based diagnostics
  - Technical computing
- GE (General Electric)
  - Diagnostics
  - Prognostics
  - Reasoning
- New York Stock Exchange
  - Business Intelligence
- Bayes Server
  - Bayesian network software
  - Technical director



Bayesian networks

- Probabilistic
- Graphical
- Not a black box
- Handle conflicting evidence
  - Unlike many rule based systems
- Multivariate
- Data driven and/or expert driven
- Missing data

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Tasks & Models

Tasks

- Classification
- Regression
- Clustering / Mixture models
- Density estimation
- Time series prediction
- Anomaly detection
- Decision Support
- Multivariate models
- Learning with missing data
- Probabilistic reasoning
- Text analytics

Models

- Multivariate Linear Regression
- Mixture models
- Time Series models
  - AR, Vector AR
- Hidden Markov Models
- Linear Kalman Filters
- Probabilistic PCA
- Factor Analysis
- Hybrid models
  - E.g. Mixtures of PPCA

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

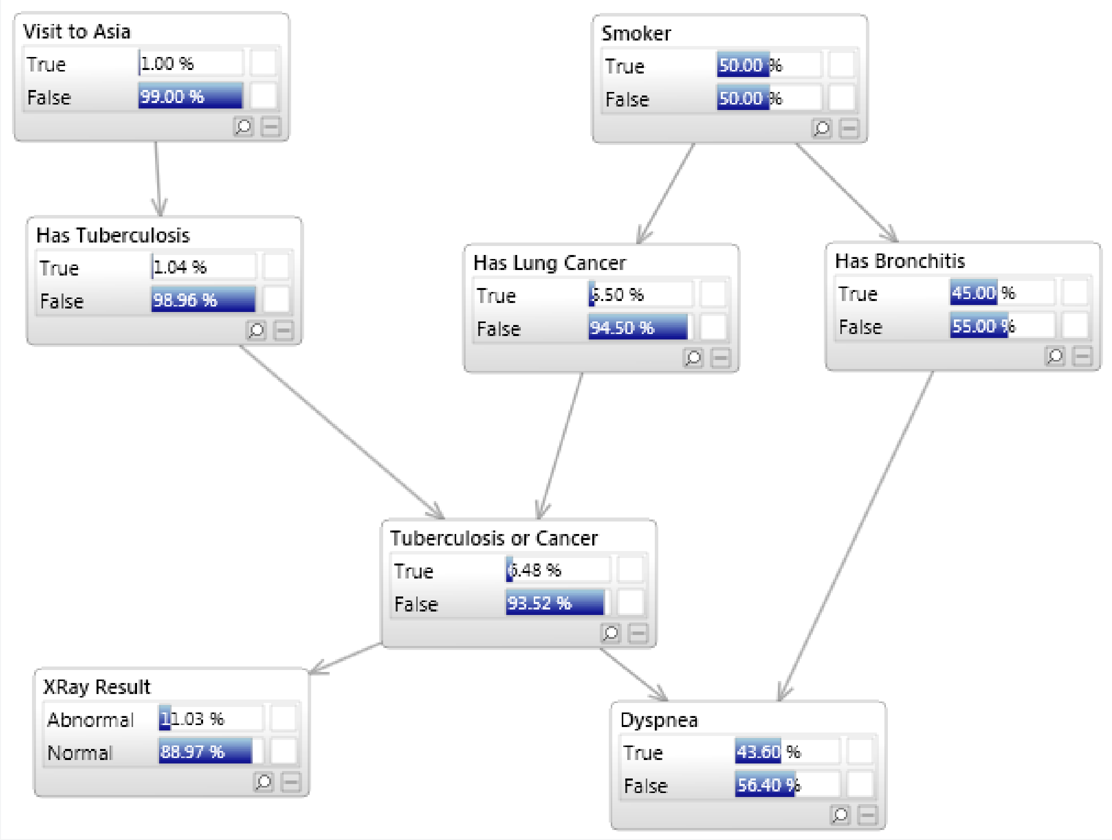
Bayesian networks

- High dimensional data
  - Humans find difficult to interpret
- Discrete and continuous variables
- Allow missing data
  - Learning
  - Prediction
- Temporal and non temporal variables in the same model

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

What are Bayesian networks?

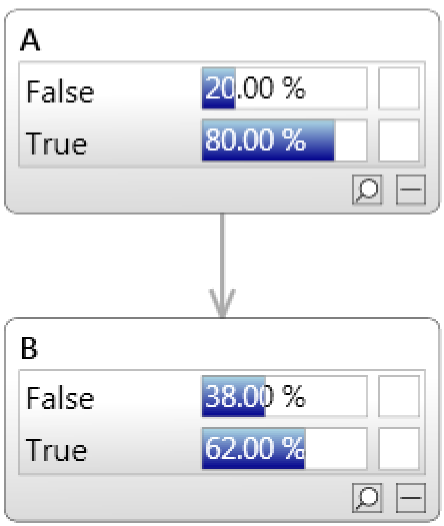
- Network with nodes and links
- Nodes represent one or more variables
- Directed links used when nodes directly influence each other
  - N.B. nodes may influence each other indirectly via other nodes
- Encode conditional independence assumptions



Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Model parameters

- Each node requires a probability distribution conditioned on its parents



A=False	A=True
0.2	0.8

A	B=False	B=True
False	0.3	0.7
True	0.4	0.6

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

What is inference?

- Asking a question, given what we know
  - E.g. $P(A|B=True, E=False)$
- We could multiply all node distributions together and get the joint distribution over all variables.
- However we can perform inference much more efficiently using the factored form

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

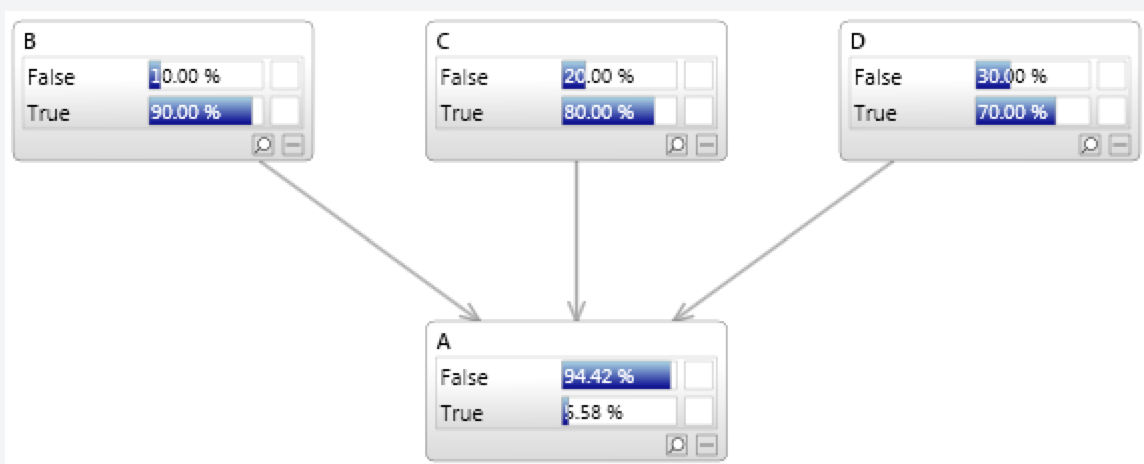
Construction

1. Add nodes (variables)
   - Manually (expert opinion)
   - From data
     - Data can be discretized if required
2. Add links
   - Manually (expert opinion)
   - From data
     - Constraint based
     - Search & score
3. Specify the parameters of the model
   - Manually (expert opinion)
   - From data
     - EM learning (handles missing data)

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Demonstration - construction

A	B	C	D
False	True	True	True
False	True	True	True
False	True	True	False
…			



Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Notes on construction

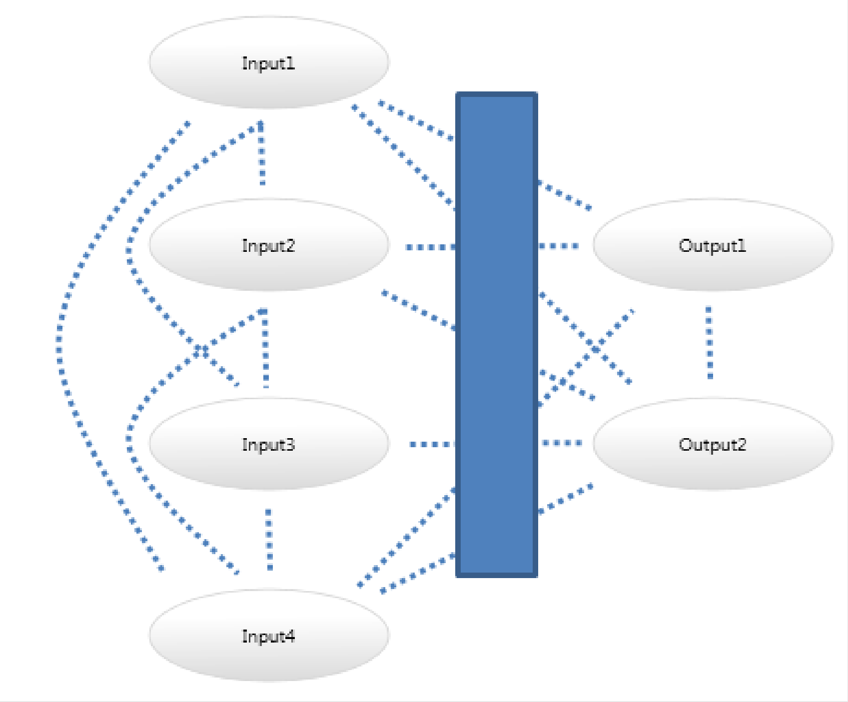
- Support for discrete & continuous
- Missing data
- Time series data

In this section we discuss classification and regression with Bayesian networks.

CLASSIFICATION / REGRESSION

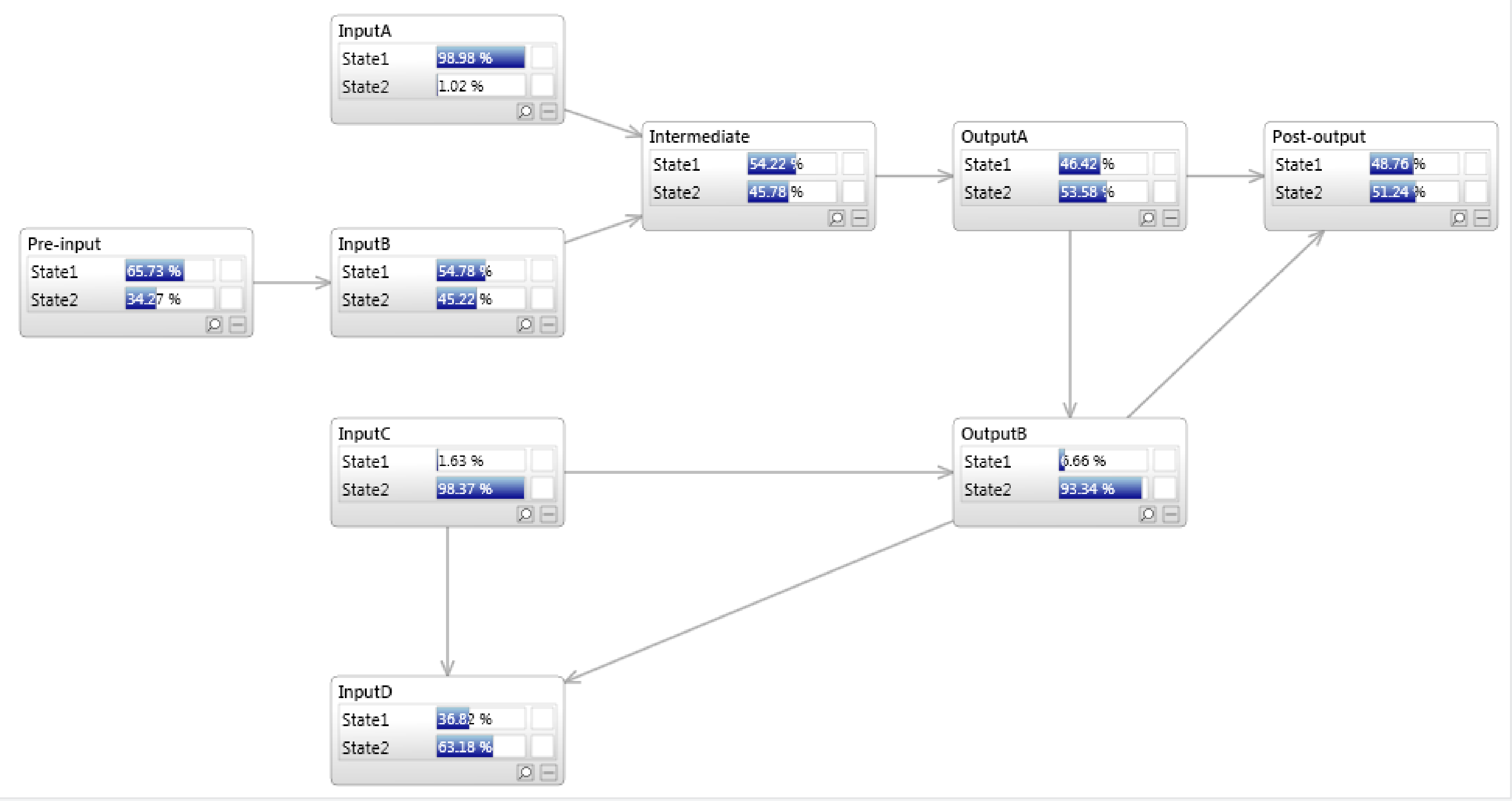
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

What is classification / regression?



- Predict unknown values (output variables), using a number of known values (input variables).
- Learning is supervised
- Classification
  - Predicting discrete variables.
- Regression
  - Predicting continuous variables.
- Examples
  - Predict the probability of a disease given symptoms
  - Predict Bull/Bear market from market indicators

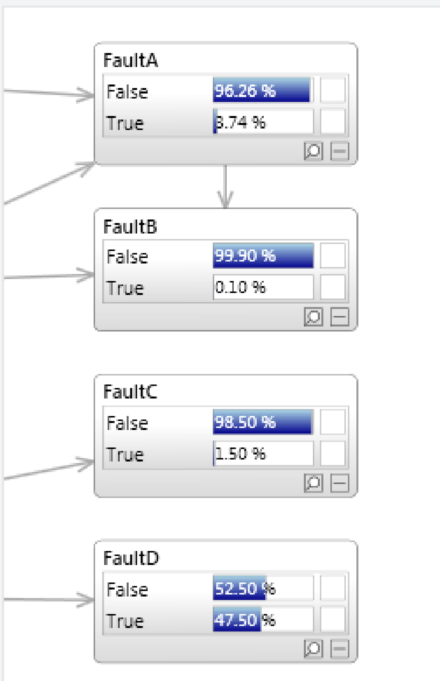
Classification structure



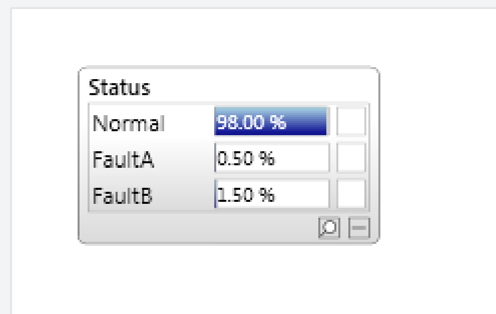
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Classification outputs

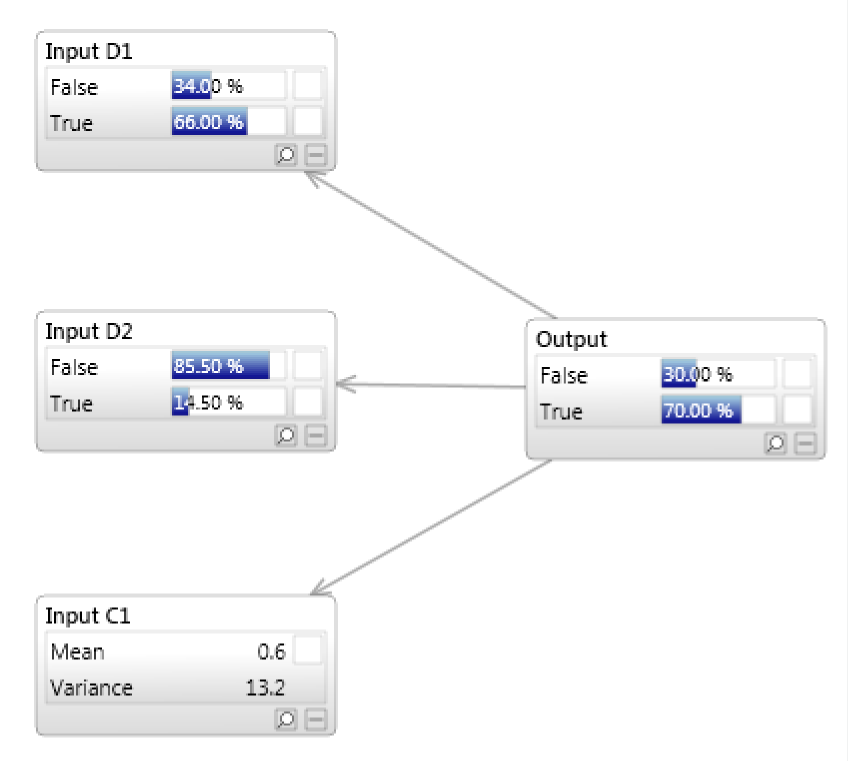
Multiple outputs



Mutually exclusive



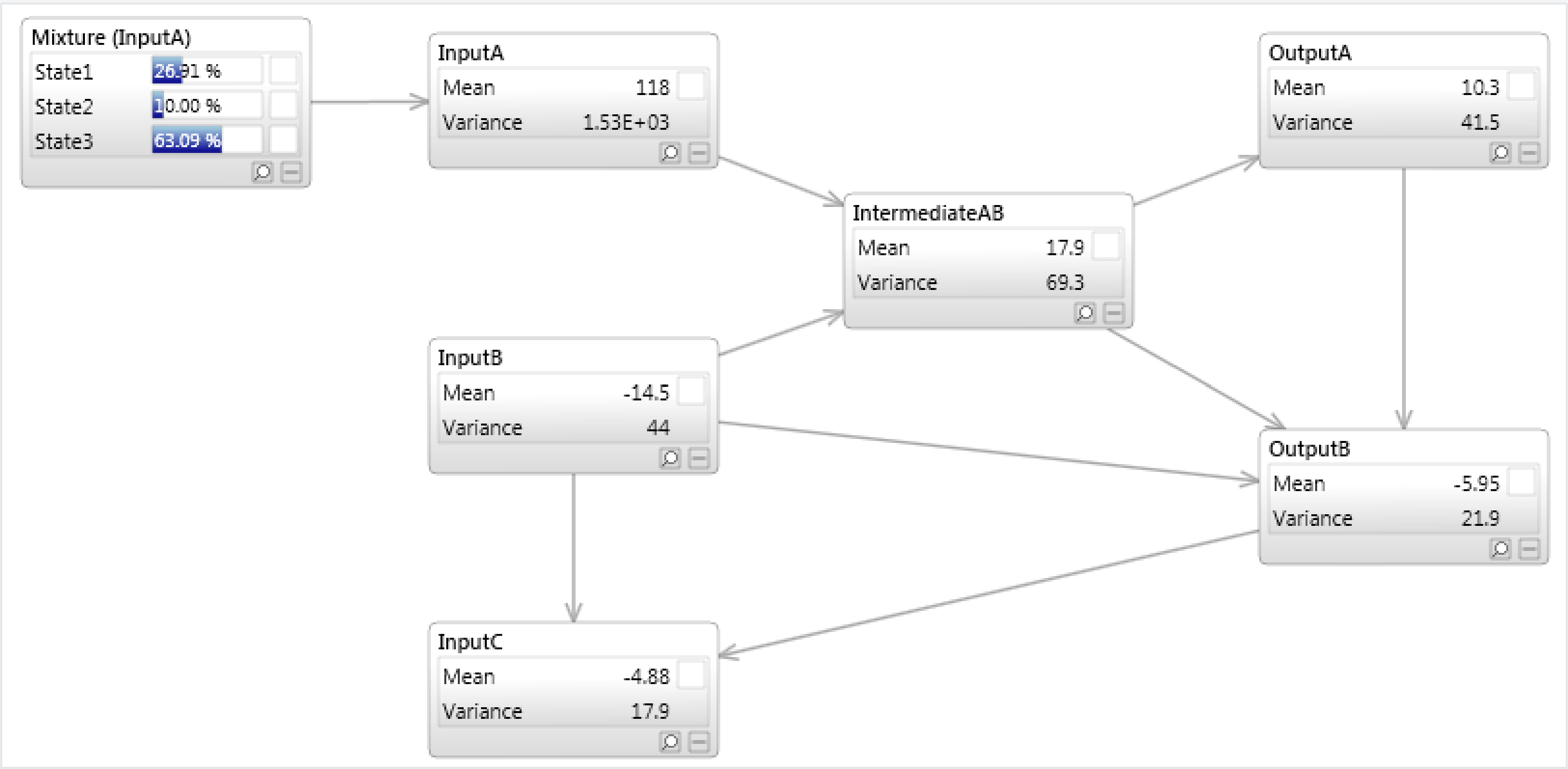
Naïve Bayes classifier



- Simple
- Fast
- Conditionally independent inputs
- Spam filters

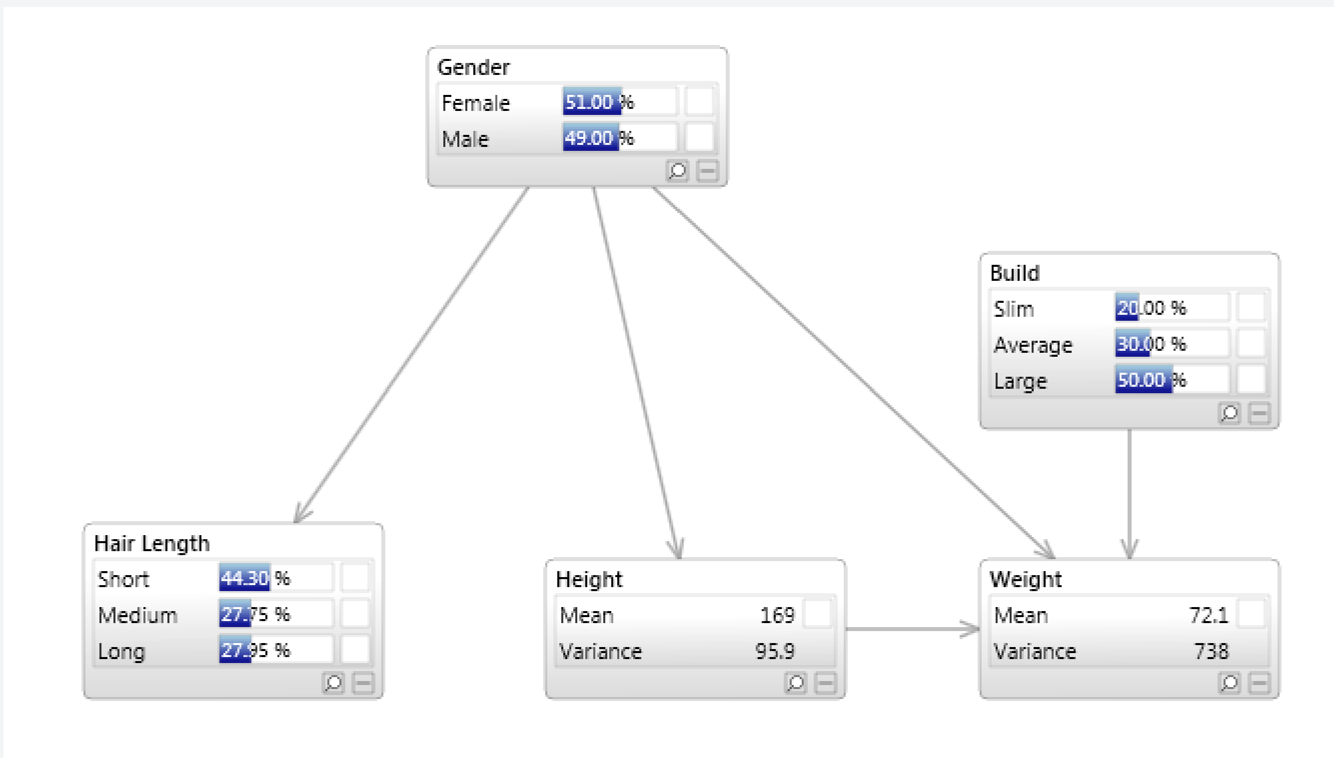
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Regression structure



Demonstration

Identification network



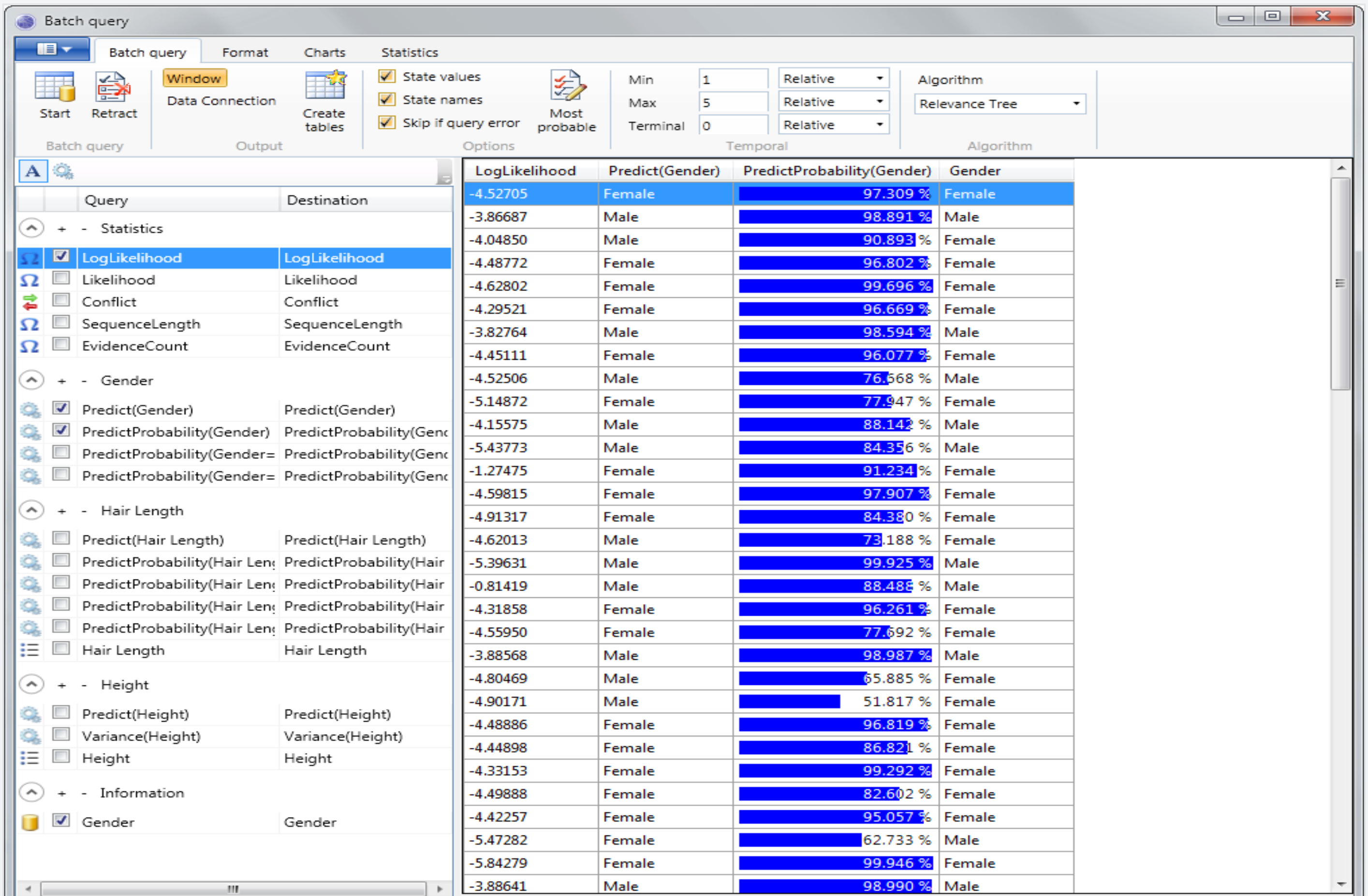
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Training

Build	Gender	Hair Length	Height	Weight
Average	Male	Short	172.35	65.57
Average	Female	Medium	155.15	49.89
?	Male	Medium	188.20	113.62
Average	Male	Short	166.47	52.88
…				

- Missing data
- Training data includes 'Gender'
- Can mix expert opinion and data

Prediction


Batch query
Batch query
Format
Charts
Statistics
Start
Retract
Batch query
Window
Data Connection
Create tables
Output
State values
State names
Skip if query error
Most probable
Options
Min 1 Relative
Max 5 Relative
Terminal 0 Relative
Temporal
Algorithm
Relevance Tree
Algorithm
Query
Destination
Statistics
LogLikelihood LogLikelihood
Likelihood Likelihood
Conflict Conflict
SequenceLength SequenceLength
EvidenceCount EvidenceCount
Gender
Predict(Gender) Predict(Gender)
PredictProbability(Gender) PredictProbability(Genc
PredictProbability(Gender= PredictProbability(Genc
PredictProbability(Gender= PredictProbability(Genc
Hair Length
Predict(Hair Length) Predict(Hair Length)
PredictProbability(Hair Len; PredictProbability(Hair
PredictProbability(Hair Len; PredictProbability(Hair
PredictProbability(Hair Len; PredictProbability(Hair
PredictProbability(Hair Len; PredictProbability(Hair
Hair Length Hair Length
Height
Predict(Height) Predict(Height)
Variance(Height) Variance(Height)
Height Height
Information
Gender Gender
LogLikelihood Predict(Gender) PredictProbability(Gender) Gender
-4.52705 Female 97.309 % Female
-3.86687 Male 98.891 % Male
-4.04850 Male 90.893 % Female
-4.48772 Female 96.802 % Female
-4.62802 Female 99.696 % Female
-4.29521 Female 96.669 % Female
-3.82764 Male 98.594 % Male
-4.45111 Female 96.077 % Female
-4.52506 Male 76.568 % Male
-5.14872 Female 77.947 % Female
-4.15575 Male 88.142 % Male
-5.43773 Male 84.356 % Male
-1.27475 Female 91.234 % Female
-4.59815 Female 97.907 % Female
-4.91317 Female 84.380 % Female
-4.62013 Male 73.188 % Female
-5.39631 Male 99.925 % Male
-0.81419 Male 88.488 % Male
-4.31858 Female 96.261 % Female
-4.55950 Female 77.692 % Female
-3.88568 Male 98.987 % Male
-4.80469 Male 65.885 % Female
-4.90171 Male 51.817 % Female
-4.48886 Female 96.819 % Female
-4.44898 Female 86.821 % Female
-4.33153 Female 99.292 % Female
-4.49888 Female 82.602 % Female
-4.42257 Female 95.057 % Female
-5.47282 Female 62.733 % Male
-5.84279 Female 99.946 % Female
-3.88641 Male 98.990 % Male

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Model performance & comparison

Confusion matrix   Display value: Count

Actual ↓	Female (Predicted)	Male (Predicted)
Female (Actual)	660	32
Male (Actual)	28	644

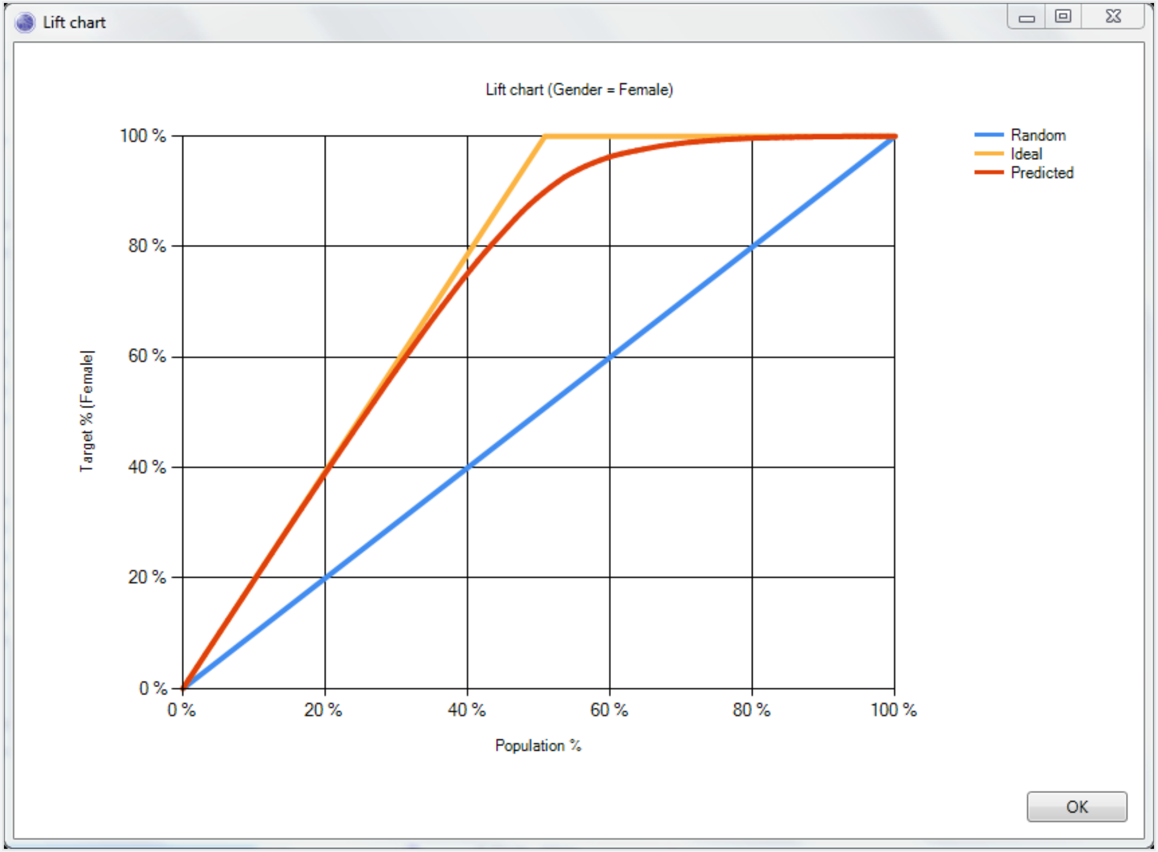
- Additional variables?
- BIC
- Confusion matrix
- Lift Chart
- Over fitting

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Confusion matrix

Actual ↓	State1 (Predicted)	State2 (Predicted)	State3 (Predicted)	State4 (Predicted)	State5 (Predicted)	State6 (Predicted)	State7 (Predicted)	State8 (Predicted)
State1 (Actual)	90.12 %	0.00 %	9.88 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
State2 (Actual)	0.00 %	100.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
State3 (Actual)	0.00 %	0.00 %	92.66 %	0.00 %	0.00 %	7.34 %	0.00 %	0.00 %
State4 (Actual)	0.00 %	0.00 %	0.00 %	100.00 %	0.00 %	0.00 %	0.00 %	0.00 %
State5 (Actual)	0.00 %	0.00 %	12.69 %	0.00 %	87.31 %	0.00 %	0.00 %	0.00 %

Lift chart


Lift chart
Lift chart (Gender = Female)
100 %
80 %
60 %
40 %
20 %
0 %
Target % (Female)
0 %
20 %
40 %
60 %
80 %
100 %
Population %
Random
Ideal
Predicted
OK

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

In this section we discuss clustering / segmentation with Bayesian networks

CLUSTERING / SEGMENTATION

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

What is clustering / segmentation?

- Unsupervised learning approach
- No outputs in the data, only inputs
- Finds natural groupings in the data
- Multivariate, handling high dimensional data
- E.g. Targeted marketing

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Usage

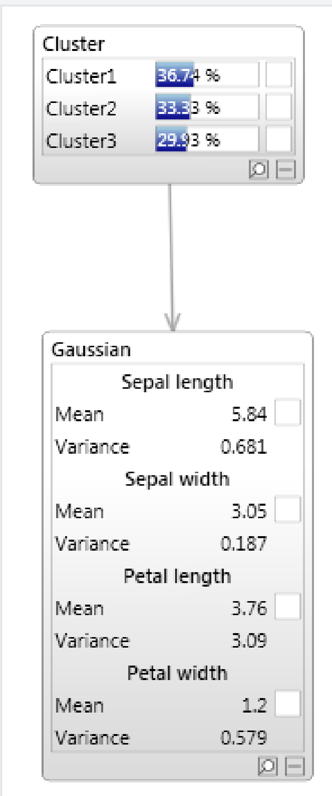
- **Data exploration**
  Mixture models are useful for identifying key characteristics of your data, such as the most common relationships between variables, and also unusual relationships.

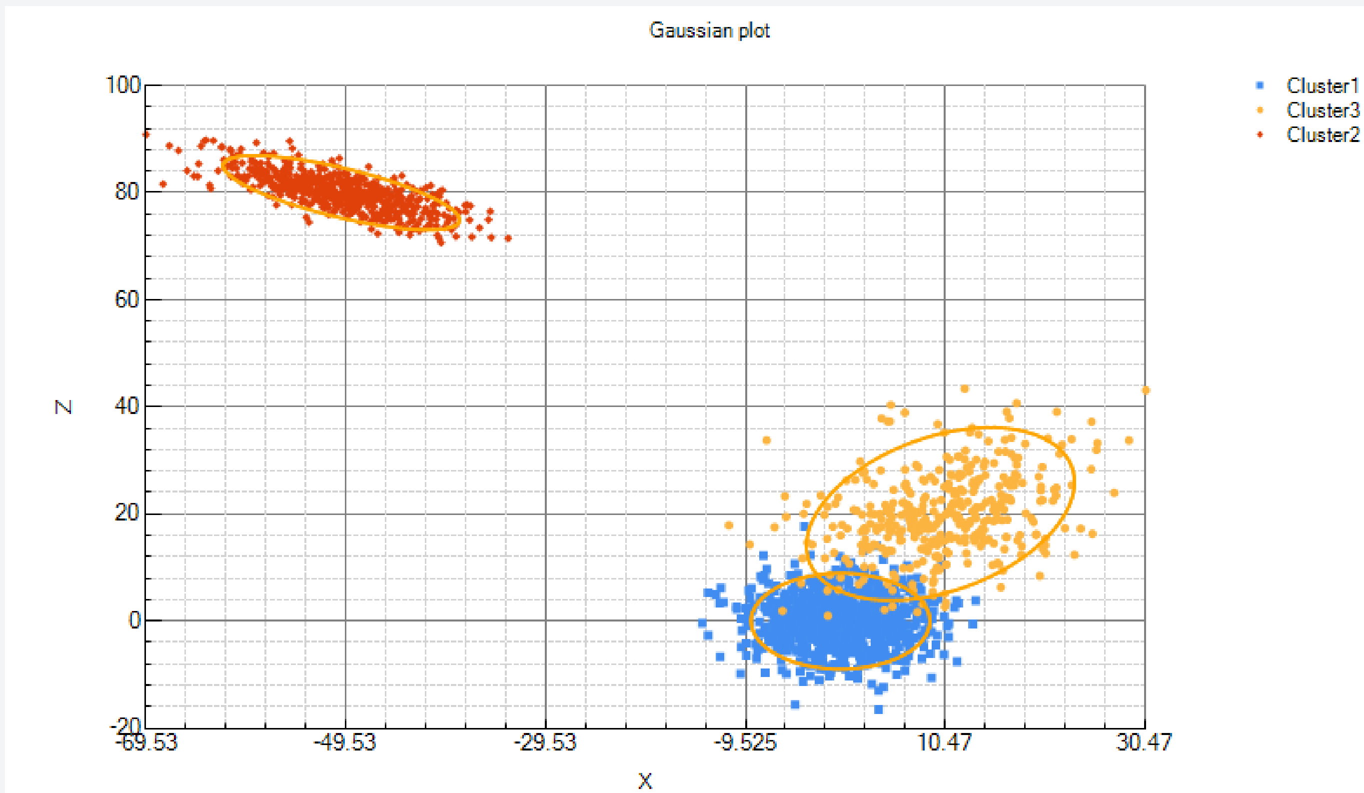
- **Segmentation**
  Because clustering detects similar groups, we can identify a group that has certain qualities and then determine segments of our data that have a high probability of belonging to that group.

- **Anomaly detection**
  Unseen data can be compared against a model, to determine how unusual (anomalous) that data is. Often the log likelihood statistic is used as a measure, as it tells you how likely it is that the model could have generated that data point. While humans are very good at interpreting 2D and 3D data, we are not so good in higher dimensional space. For example a mixture model could have tens or even hundreds of dimensions.

- **Prediction**
  Although Mixture models are an unsupervised learning technique, we can use them for prediction if during learning, we include variables we wish to predict (output variables).

Demonstration – mixture model





Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

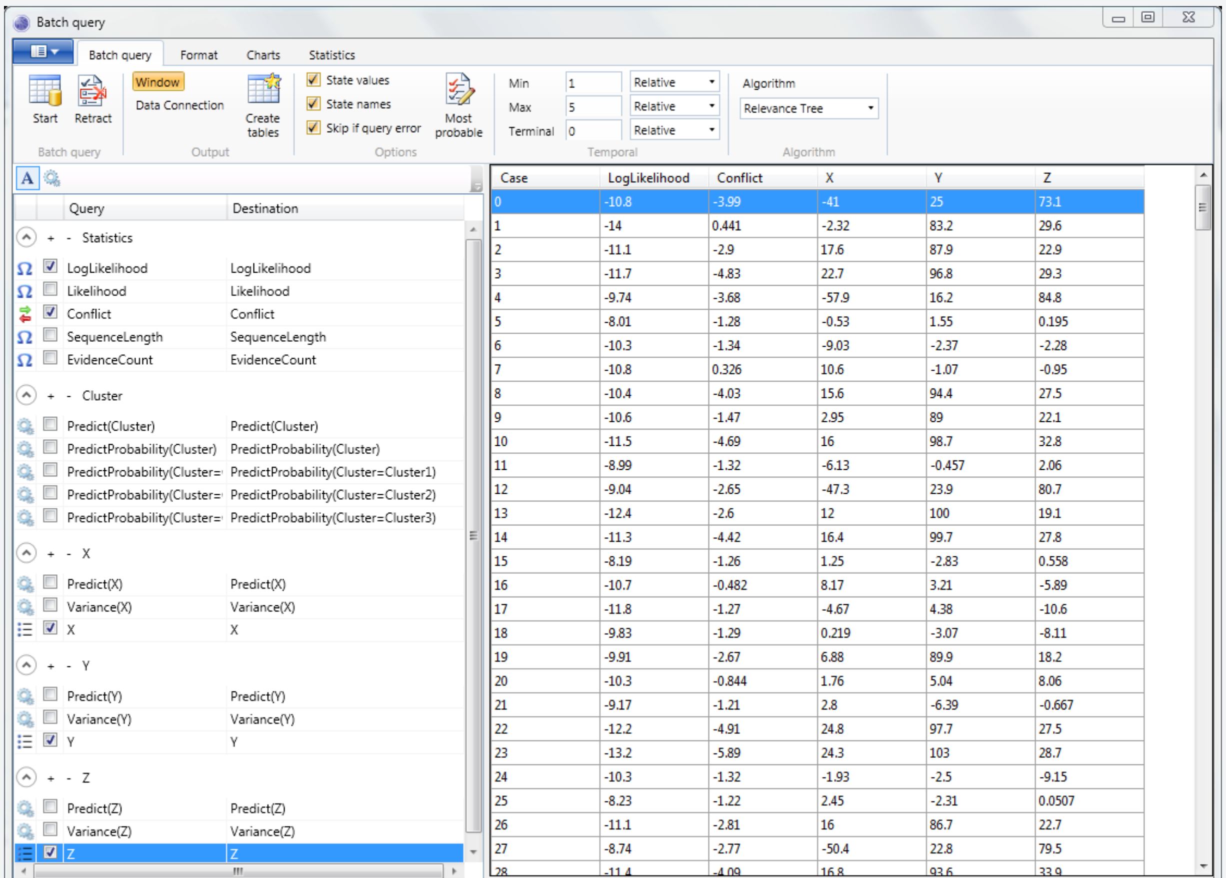
Mixture model – anomaly detection

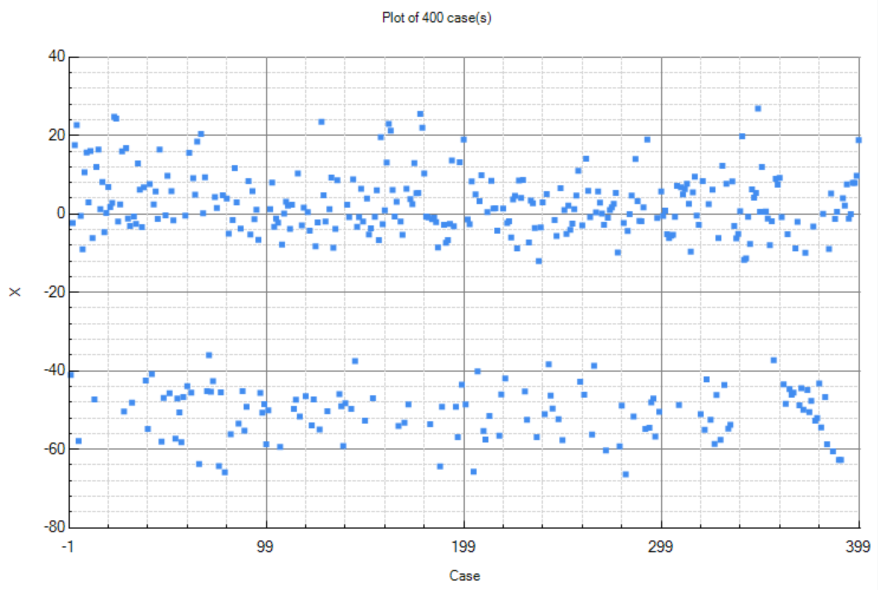
X	Y	Z
-41.04	25.04	73.13
-2.32	83.20	29.59
17.57	87.94	22.85
…		

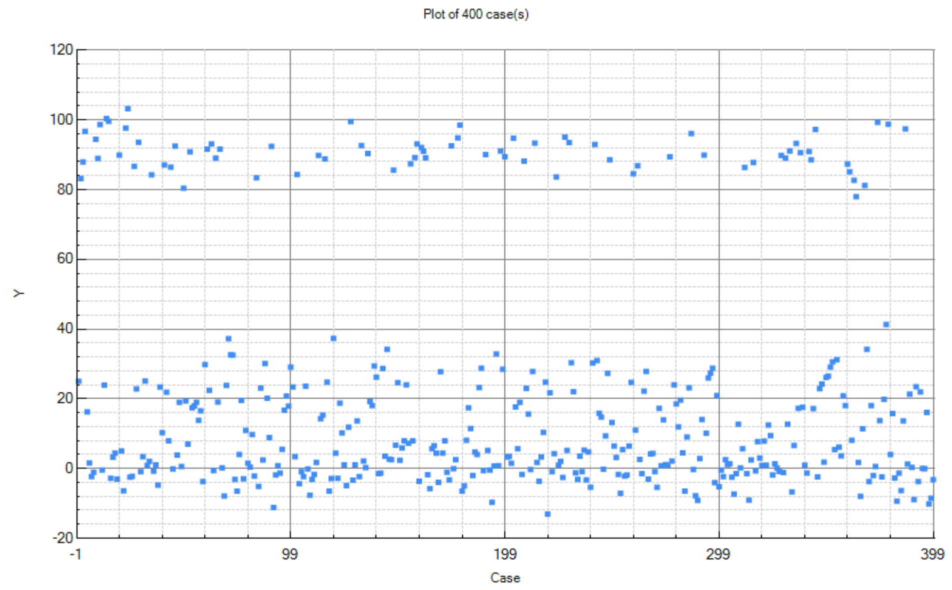
- No data mapped to Cluster variable
- Missing data allowed
- Predict (Cluster)
- Log likelihood
- Conflict

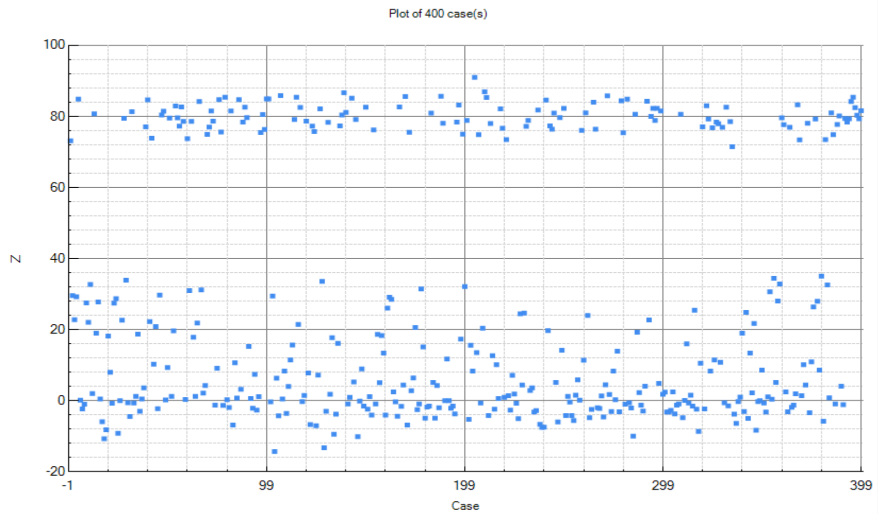
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Mixture model – batch prediction





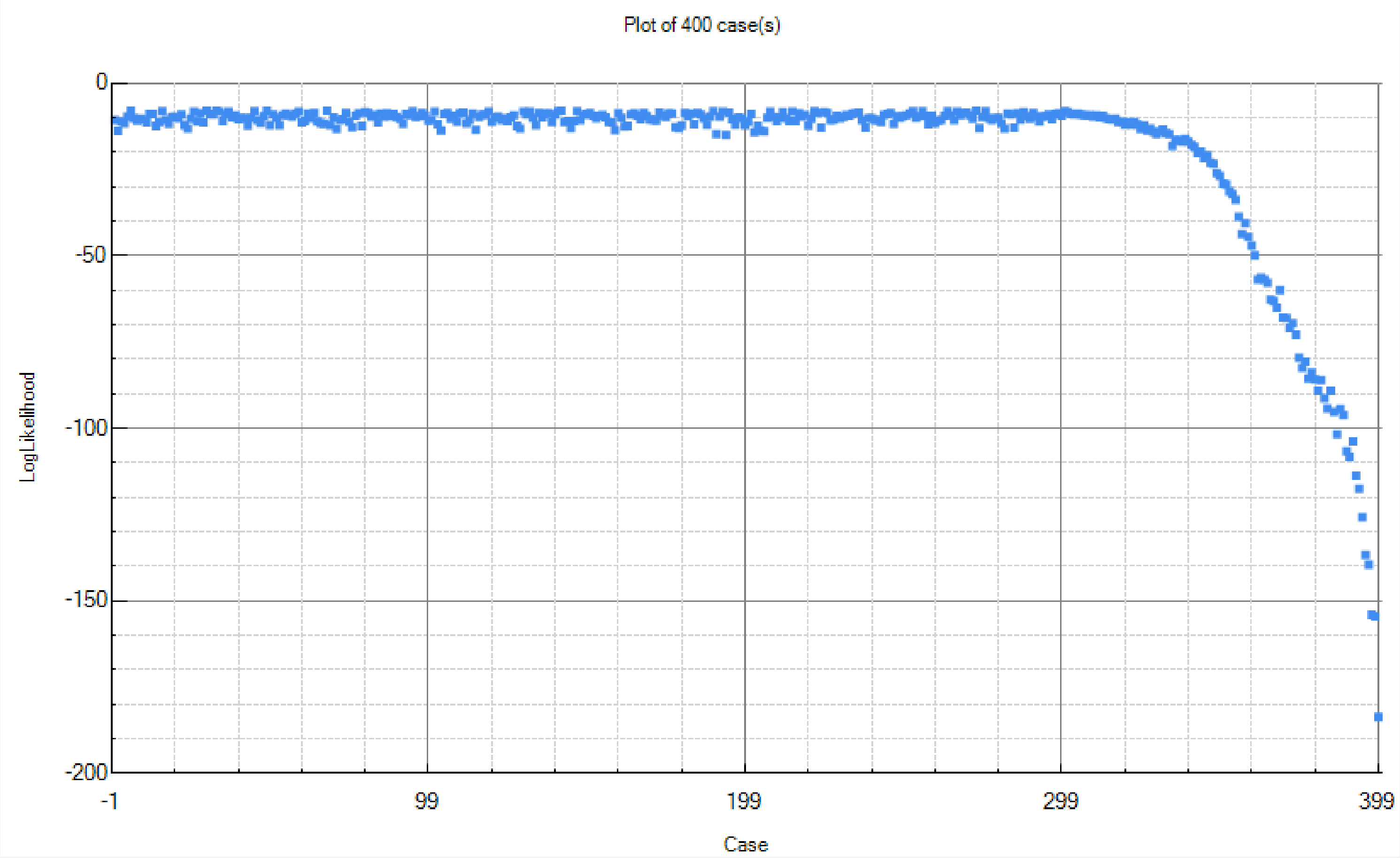




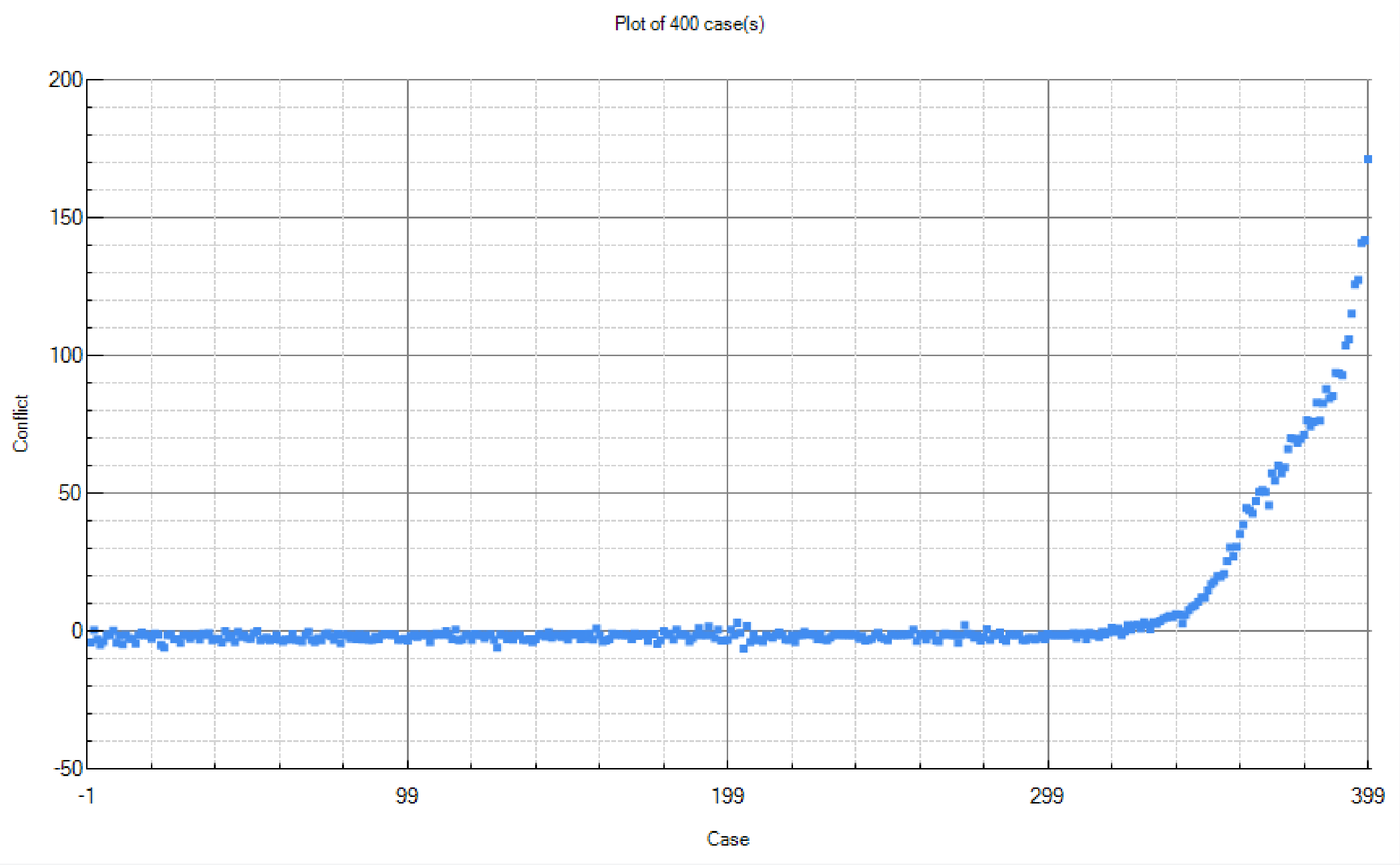
- Bi-model or tri-modal
- Univariate analysis looks normal

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Multivariate prediction (log-likelihood)



Multivariate prediction (conflict)



Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

In this section we discuss time series prediction with Bayesian networks

TIME SERIES PREDICTION

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Time series models

- Known as Dynamic Bayesian networks
- Discrete & continuous
- Multivariate time series
  - (Partial) Auto correlations
  - (Partial) Cross correlations
- A node can be linked to itself

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Time series

- Temporal & non temporal variables
- Classification, Regression, Log likelihood
- Modelling time series data without a time series model

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Case	Gender	Age
1	Female	35
2	Male	?
3	Female	25
…		

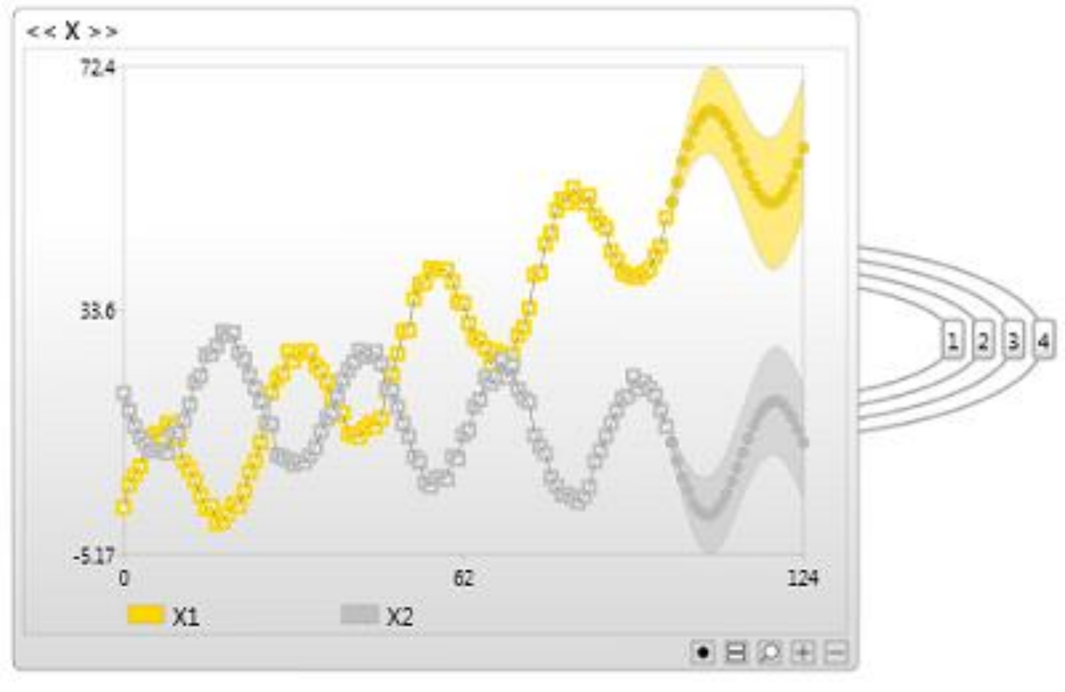
Case	Time	Transition	Obs1	Obs2
1	0	Cluster 0	12.4	15.5
1	1	Cluster 1	14.2	13.45
2	0	Cluster 1	?	8.6
2	1	Cluster 1	12.3	14.0
2	2	Cluster 1	18.3	13.5
3	0	Cluster 2	9.3	8.7
…				

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

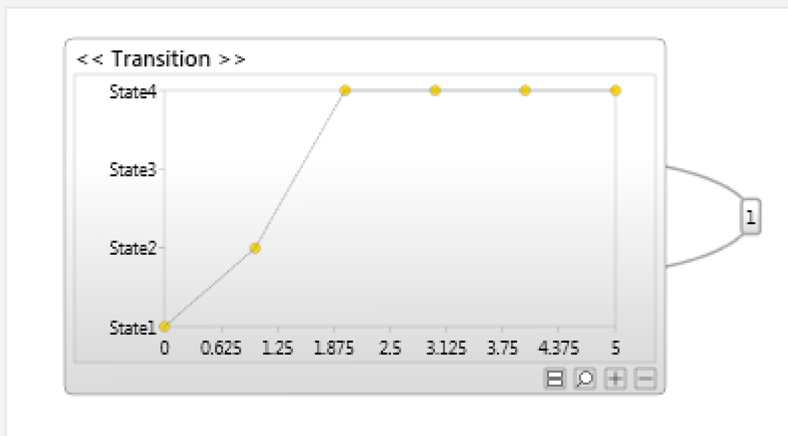
Types of time series model

- Auto regressive / vector auto regressive
- N-order markov models
- Hidden markov models
- Kalman filters
- Any of these well known models can be extended

Vector auto regressive

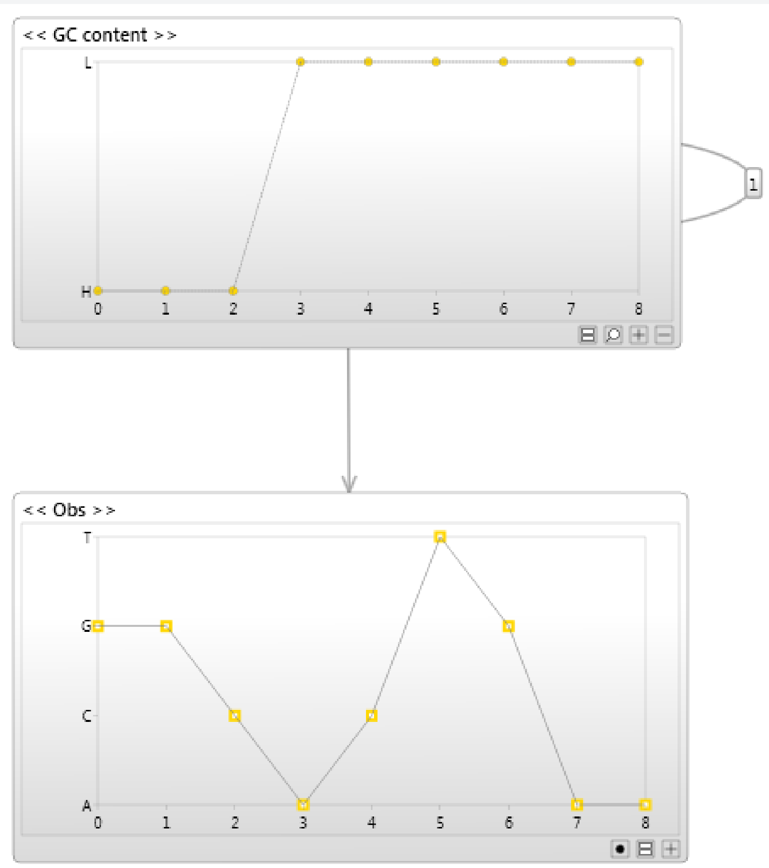


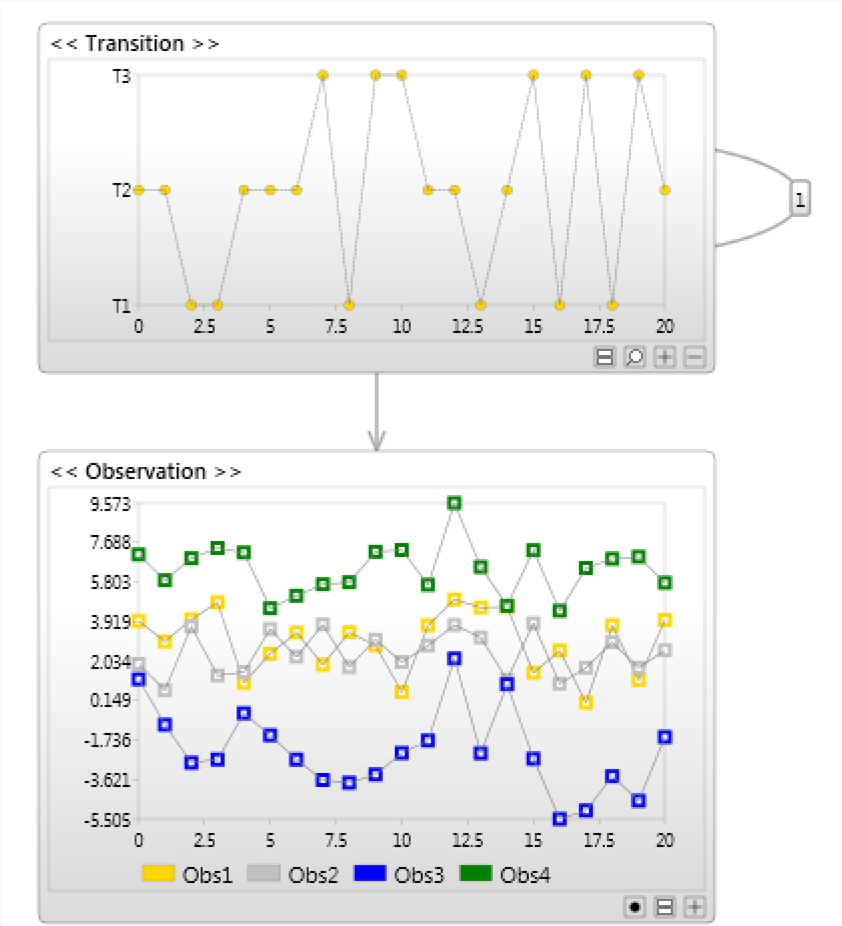
N-order markov models



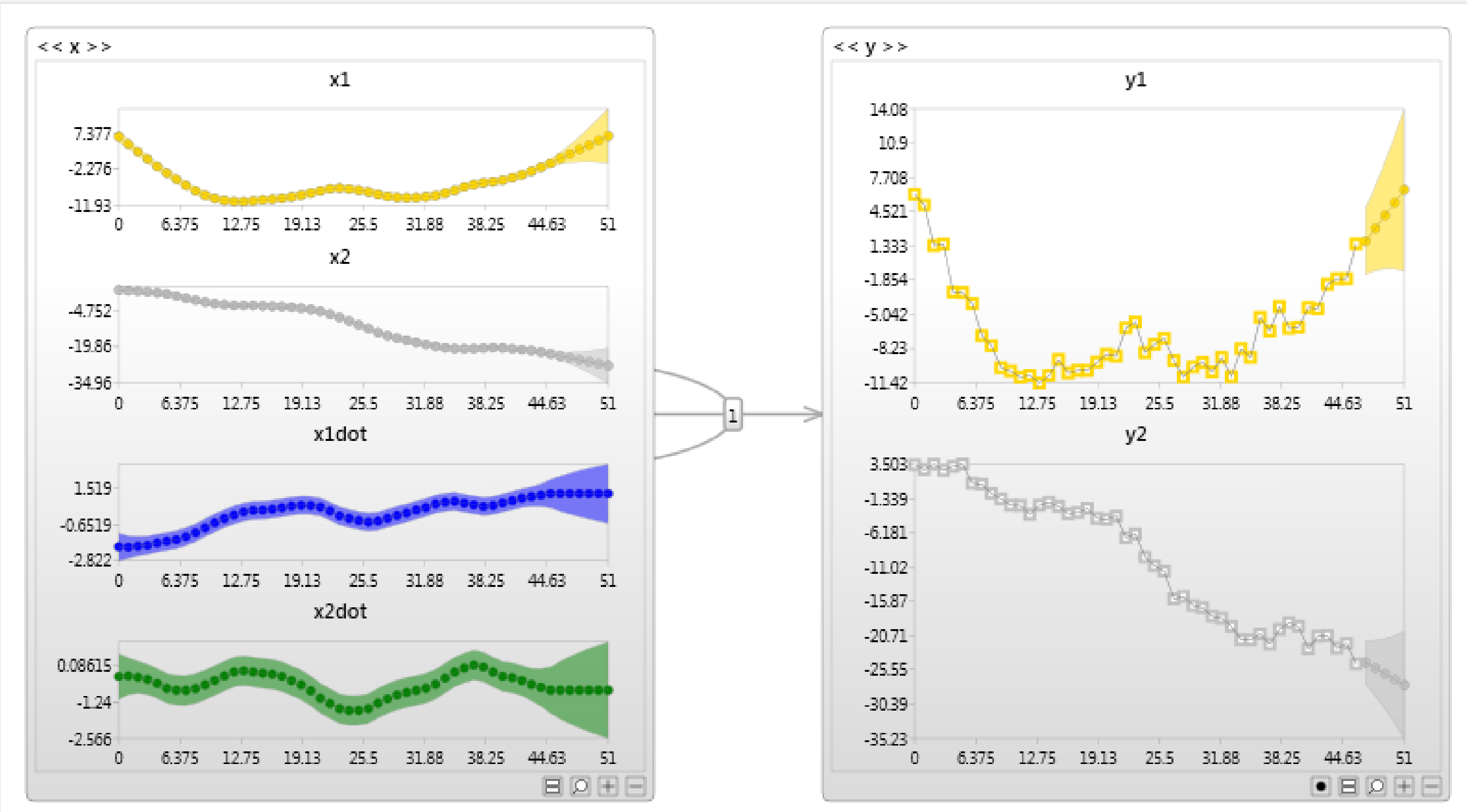
Bayes Server
intelligent systems specialists
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Hidden Markov model



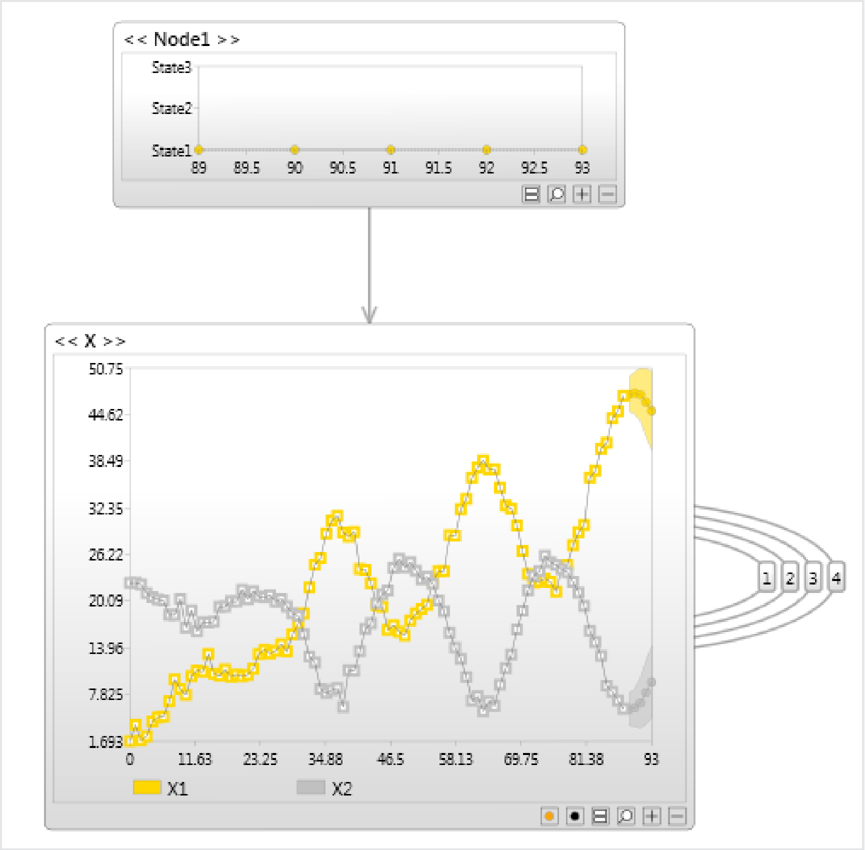


Kalman Filter



Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Mixture of vector auto regressive

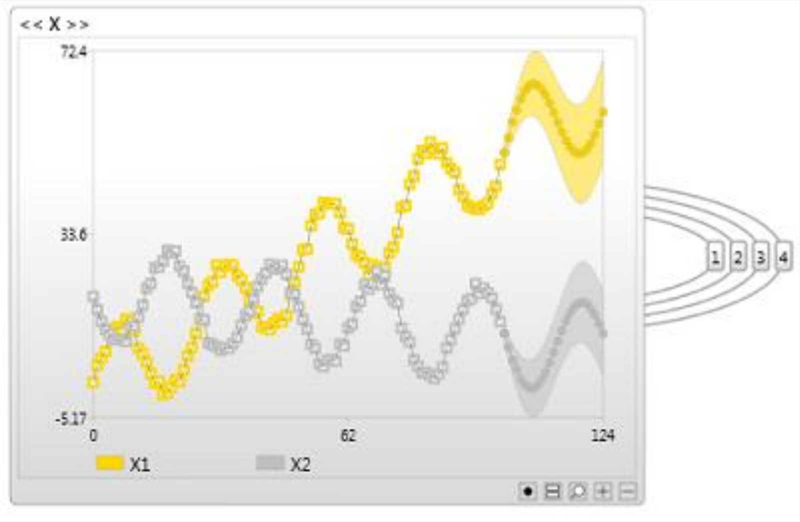


Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Types of time series prediction

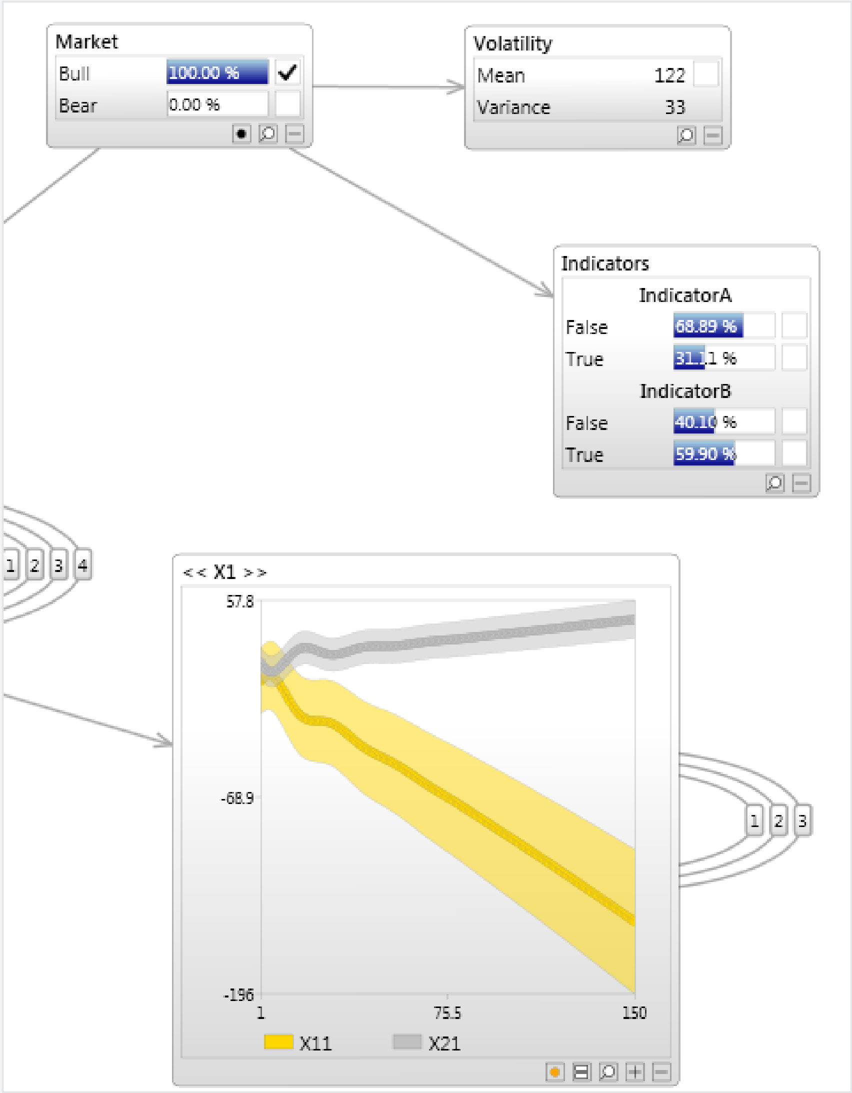
- **Prediction** - calculating queries a number of time steps into the future.
- **Filtering** - calculating queries at the current time.
- **Smoothing** - calculating queries a number of time steps into the past (calculating historic values)
- **Most probable sequence** - calculating the most likely sequence of past values (generalized version of the viterbi algorithm)

Demonstration



- Parameter Learning
- Prediction
  - Data explorer
  - Batch queries
- Sampling
  - Charting
- Structural learning
  - Determine links & orders

Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer



Questions