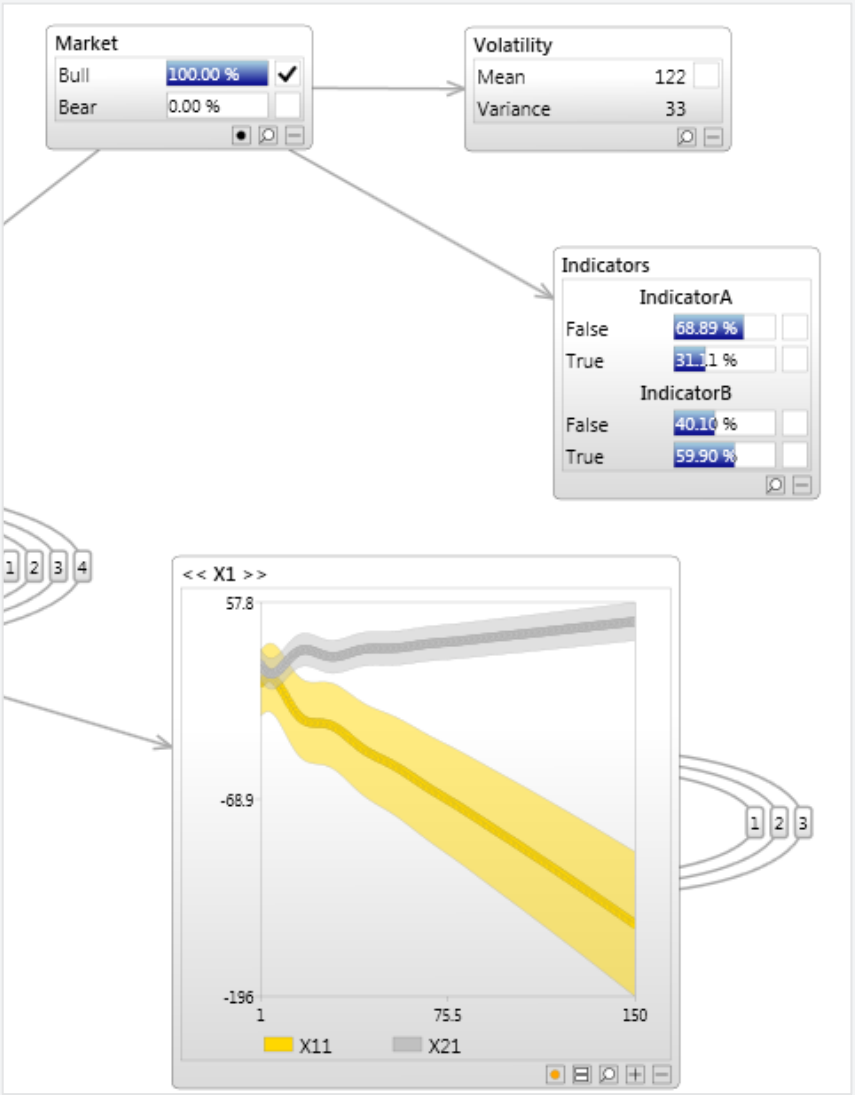
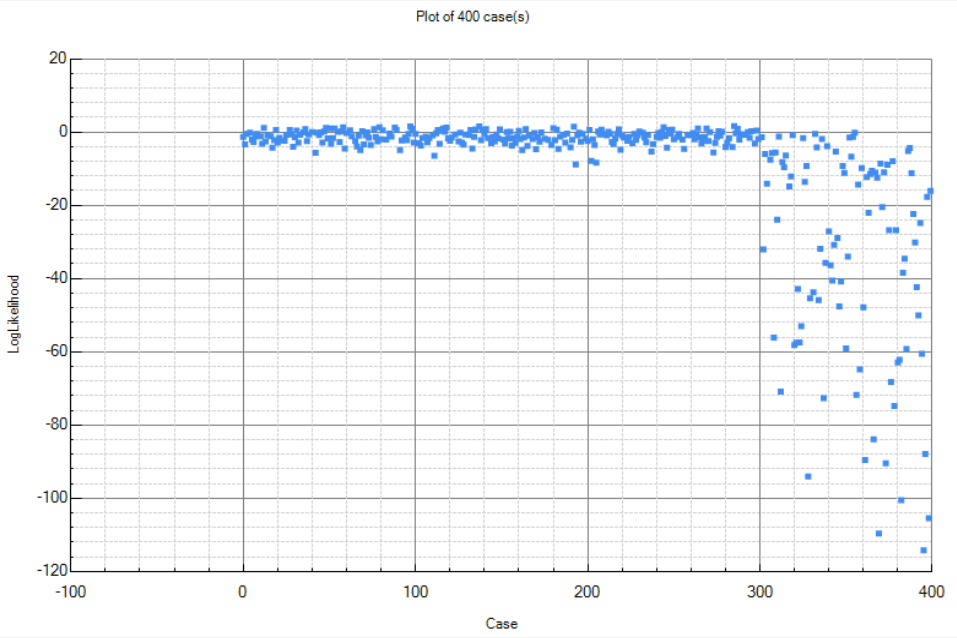
Bayes Server
intelligent systems specialists

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



Anomaly detection with Bayesian networks

John Sandiford



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

Contents

- Background
- What is anomaly detection?
- Bayesian networks
- Anomaly detection – supervised
- Anomaly detection – semi supervised
- Anomaly detection - time series

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

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

What is anomaly detection?

Anomaly detection, or outlier detection, is the process of identifying data which is unusual.

- System health monitoring
  - Advanced warning of mechanical failure
- Fault detection
  - Isolate faulty components
- Fraud detection
  - Can warn financial institutions of fraudulent transactions

- Pattern detection
  - Can detect unusual patterns
- Pre-processing
  - E.g. removal of unusual data, before building statistical models.

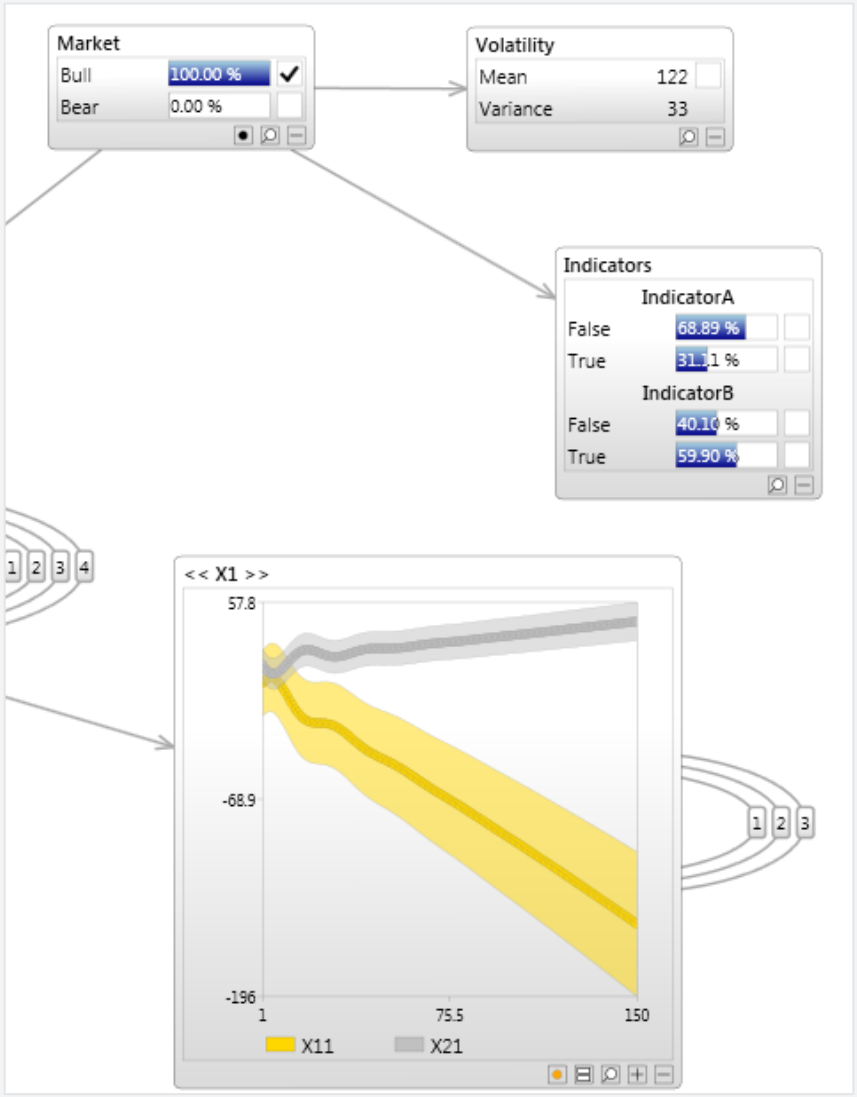
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Types of anomaly detection

- Supervised
  - Labelled data
  - Specific faults
- Semi supervised
  - 'Normal' training data
- Unsupervised
  - Normal and abnormal training data

Time Series can use all of the above

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



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

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

Anomaly detection with Bayesian networks

- High dimensional data
  - Humans find difficult to interpret
  - Anomalies may not be visible on individual variables
- Discrete and continuous variables
- Allow missing data
  - Learning
  - Prediction/anomaly detection
- Temporal and non temporal variables in the same model

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

In this section we discuss anomaly detection with Bayesian networks, using labelled data

SUPERVISED ANOMALY DETECTION

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

Comparison

Advantages

- Learning is focused
- Prediction is specific and has an associated probability
- Diagnostics easier

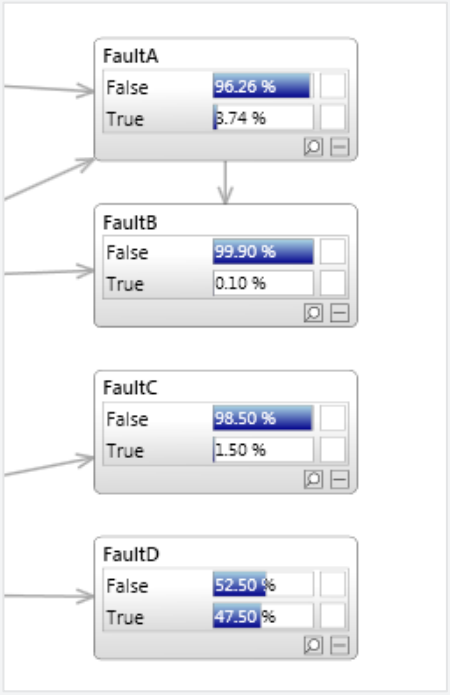
Disadvantages

- Anomalies tend to be different
  - Past anomalies may not predict future anomalies well
- Expense of labelling
  - E.g. Cost of experts
- Insufficient data labelled anomalous.
- It is too difficult to manually identify anomalous data.
  - Perhaps because the data is high dimensional, or is a complex time series or both.

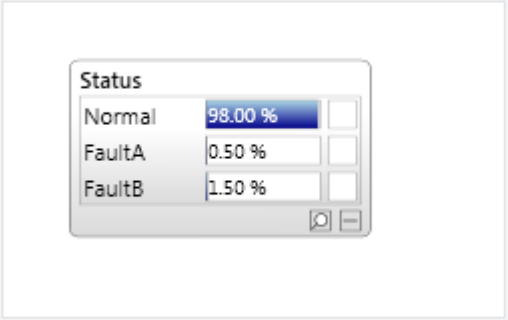
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Classification

Multiple outputs



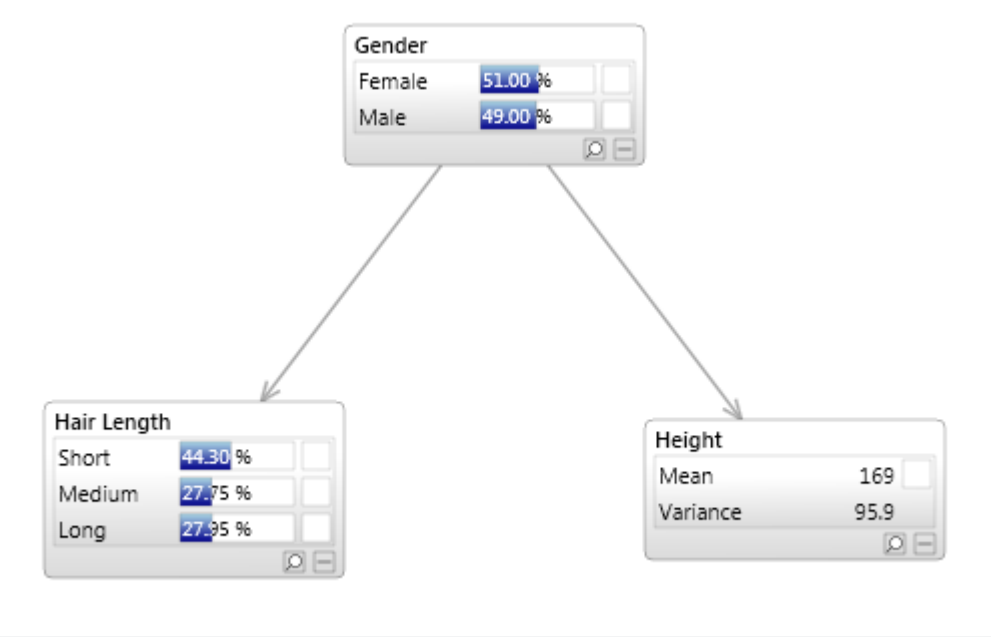
Mutually exclusive



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

Demonstration

Identification network



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

Training

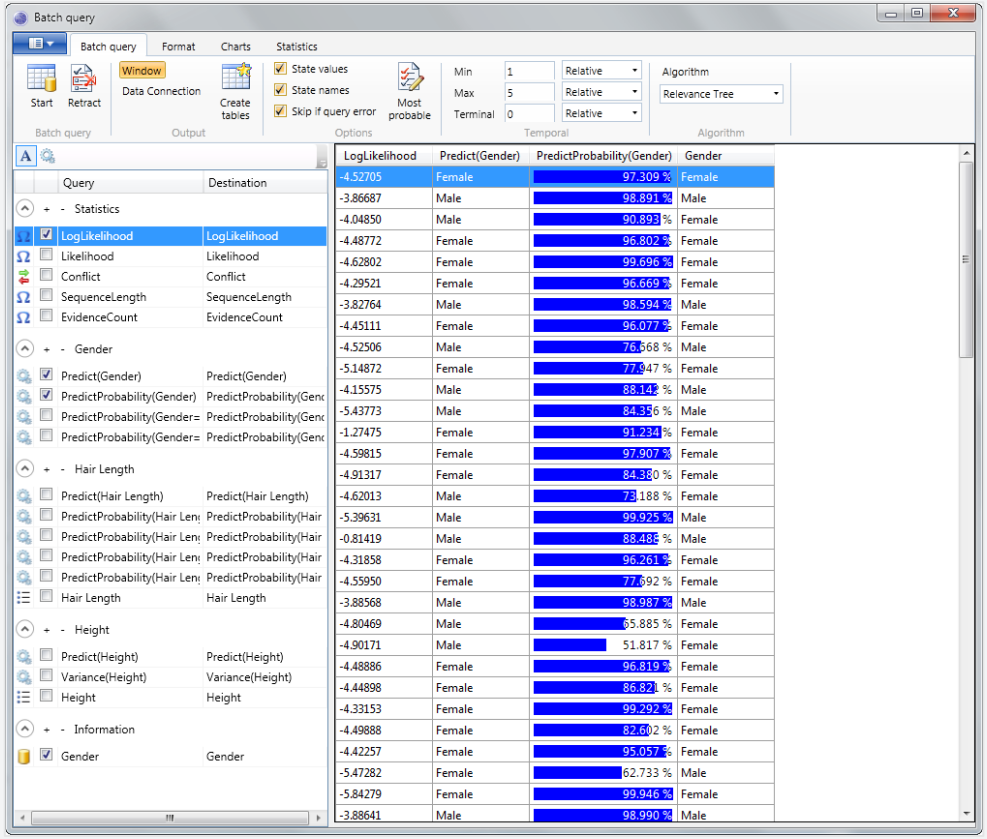
Case	Gender	Hair Length	Height
0	Male	Short	165.2878011
1	Female	Medium	172.7821406
2	Female	Medium	167.0237128
3	Female	Long	154.2981223
4	Female	Medium	160.7306941
5	Male	Short	168.3076565
6	Male	Medium	
7	Male	Medium	176.7689419
8	Male	Short	159.9983284
9	Female	Medium	177.5093331
10	Female	Long	154.1897993
11	Male	Short	179.6507421
12	Female	Medium	174.7971554
13	Female	Long	164.3262306
14	Male	Short	190.0837395
15	Male	Short	171.7654745
16	Female	Medium	158.0243525
17	Female	Long	164.3285404
18	Female	Medium	161.061252
19	Female	Medium	
20	Male	Short	189.3959879
21	Female	Short	152.7301208
22	Female	Long	165.950998
23	Male	Medium	177.2188164
...	...	...	...

- Expert opinion
- Learn parameters from data

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

Prediction

Case	Gender	Hair Length	Height
1	Female	Medium	159.64532
2	Male	Short	178.50209
3	Female	Short	170.2725
4	Female	Medium	160.31395
5	Female	Long	156.32858
6	Female	Long	165.43799
7	Male	Short	177.59889
8	Female	Medium	161.11003
9	Male	Short	166.09811
10	Female	Long	173.34889
11	Male	Short	169.16522
12	Male	Medium	179.45741
13	Female	Long	
14	Female	Medium	158.67832
15	Female	Long	171.75507
16	Female	Short	165.4013
17	Male	Short	188.6639
18	Male	Short	
19	Female	Long	165.88785
20	Female	Medium	168.43815
21	Male	Short	178.84286
22	Female	Short	164.10128
23	Female	Medium	173.39975
...	...	...	...



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

Model performance & comparison

Actual ↓	Female (Predicted)	Male (Predicted)
Female (Actual)	42	7
Male (Actual)	7	38

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

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

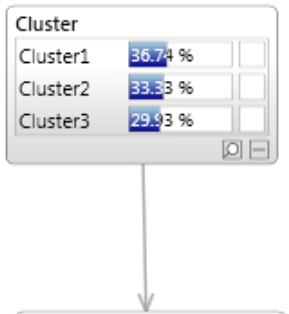
In this section we discuss anomaly detection with Bayesian networks, using 'normal' training data.

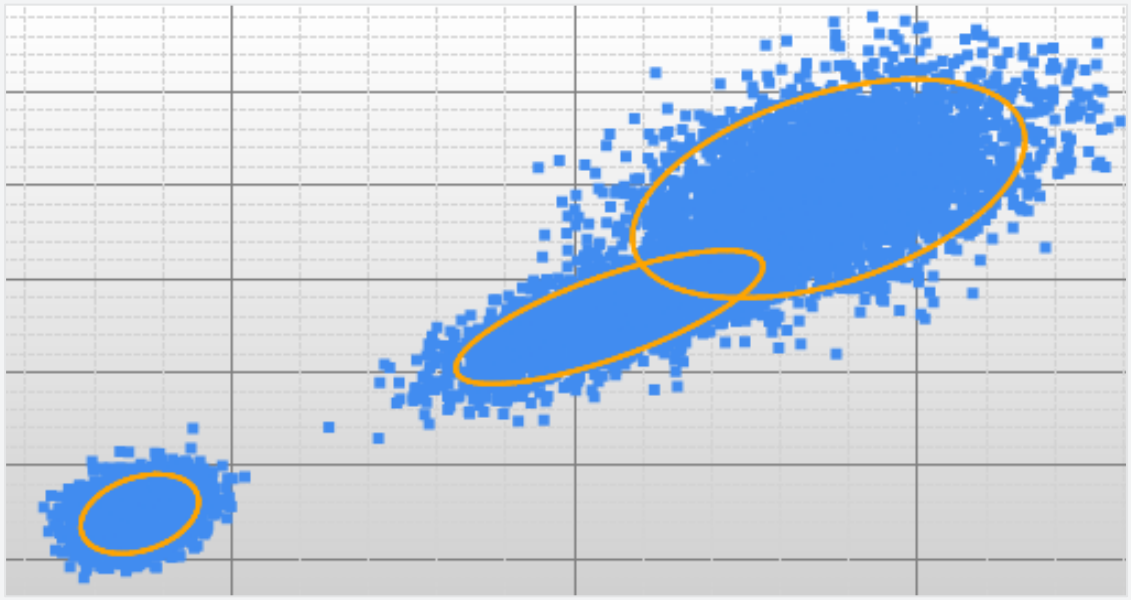
SEMI SUPERVISED ANOMALY DETECTION

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

Demonstration

Mixture model





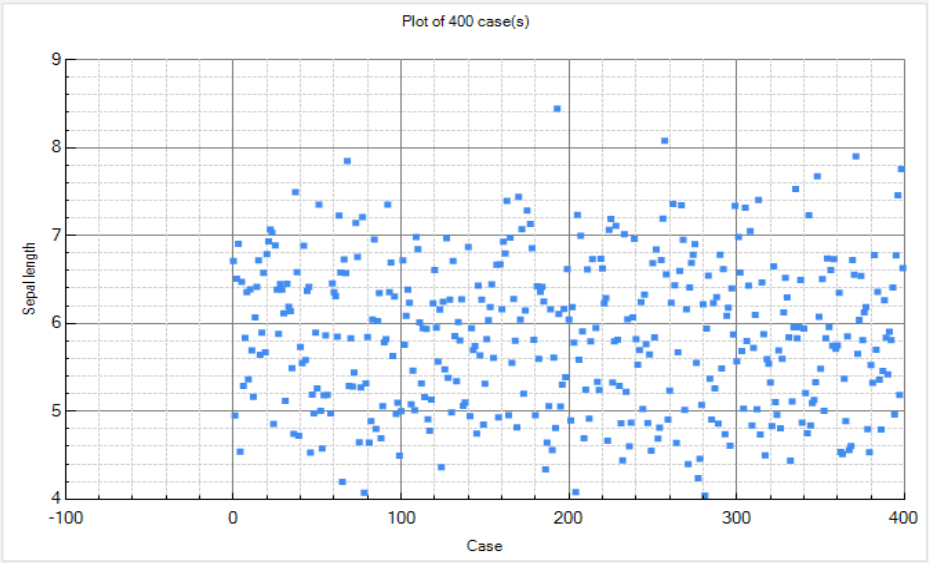
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

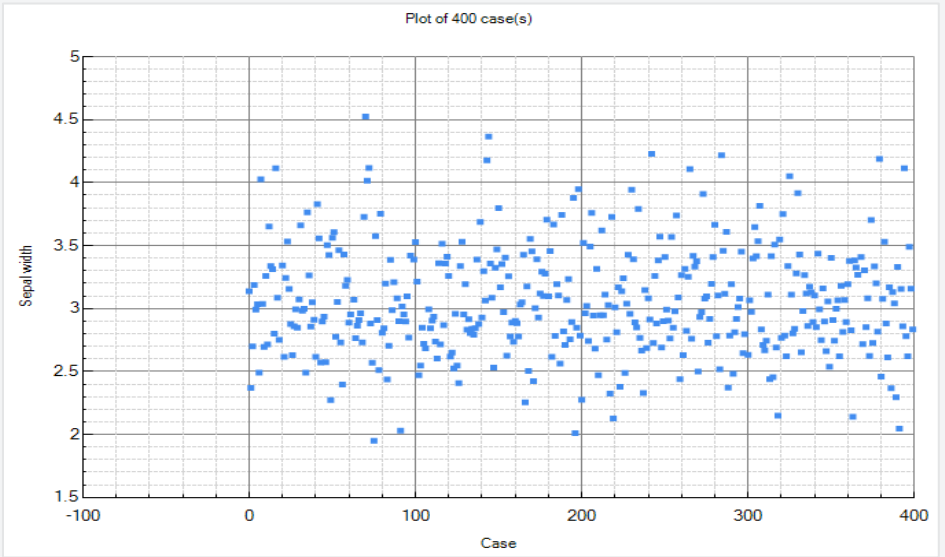
Prediction

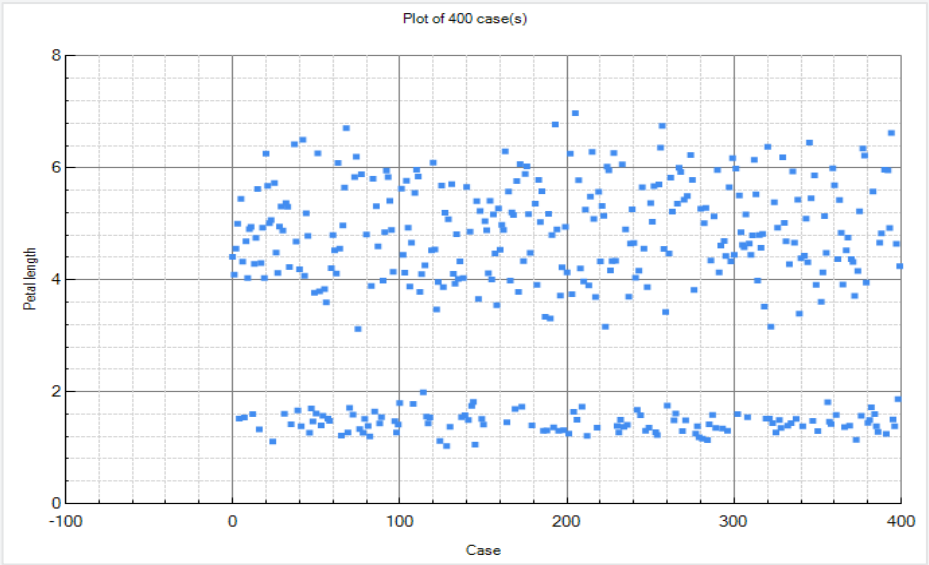
Case	Petal length	Petal width	Sepal length	Sepal width
171	3.78543176	1.138440126	6.041902403	2.422836148
172	6.072084389	2.503914705	7.072203848	3.002792376
173	1.730709603	0.379055624	5.198587724	3.390378817
174	4.337948261	1.241569329	6.145304757	2.92751191
175	5.893662725	1.804904604	7.284342906	3.11984017
176	6.030745392	2.485271154	7.295575651	3.293069842
177	5.172481556	1.515917417	7.131254692	3.098449165
178	4.480773312	1.470831396	6.855488491	3.276610303
179	1.396363339	0.129860902	5.812072163	3.704225375
180	1.399720641	0.209082919	4.955190155	3.09697432
181	5.360363857	2.03599278	6.423200941	3.456997381
182	3.908567779	1.199880583	5.595978365	2.616166089
183	5.786045951	2.228005652	6.360549274	3.666645536
184	5.035588584	1.684820935	6.41691616	2.782970144
185	5.582729759	2.14241093	6.25069291	3.193143013
186	1.301619104	0.124755369	4.338984258	3.102718803
...	...	...	...	...

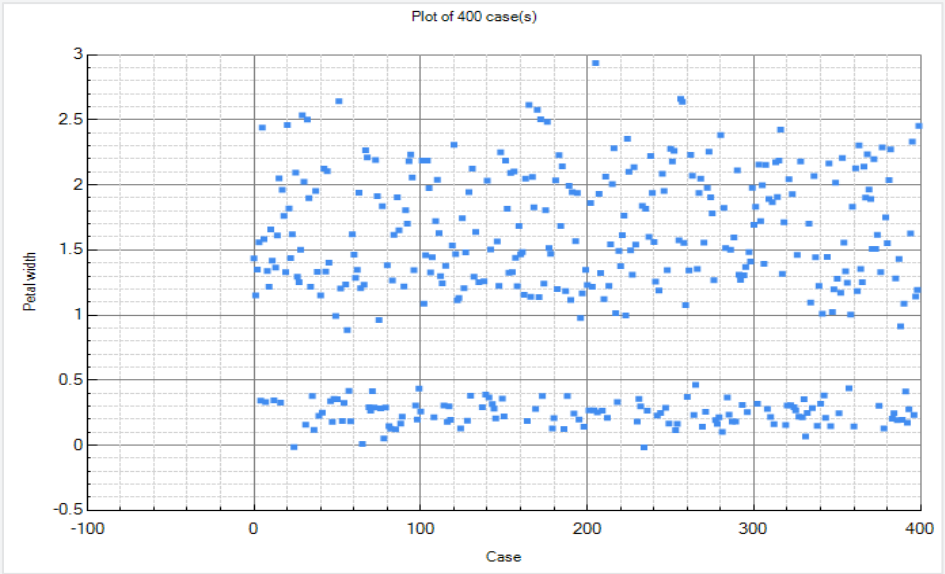
- No data mapped to Cluster variable
- Missing data allowed

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



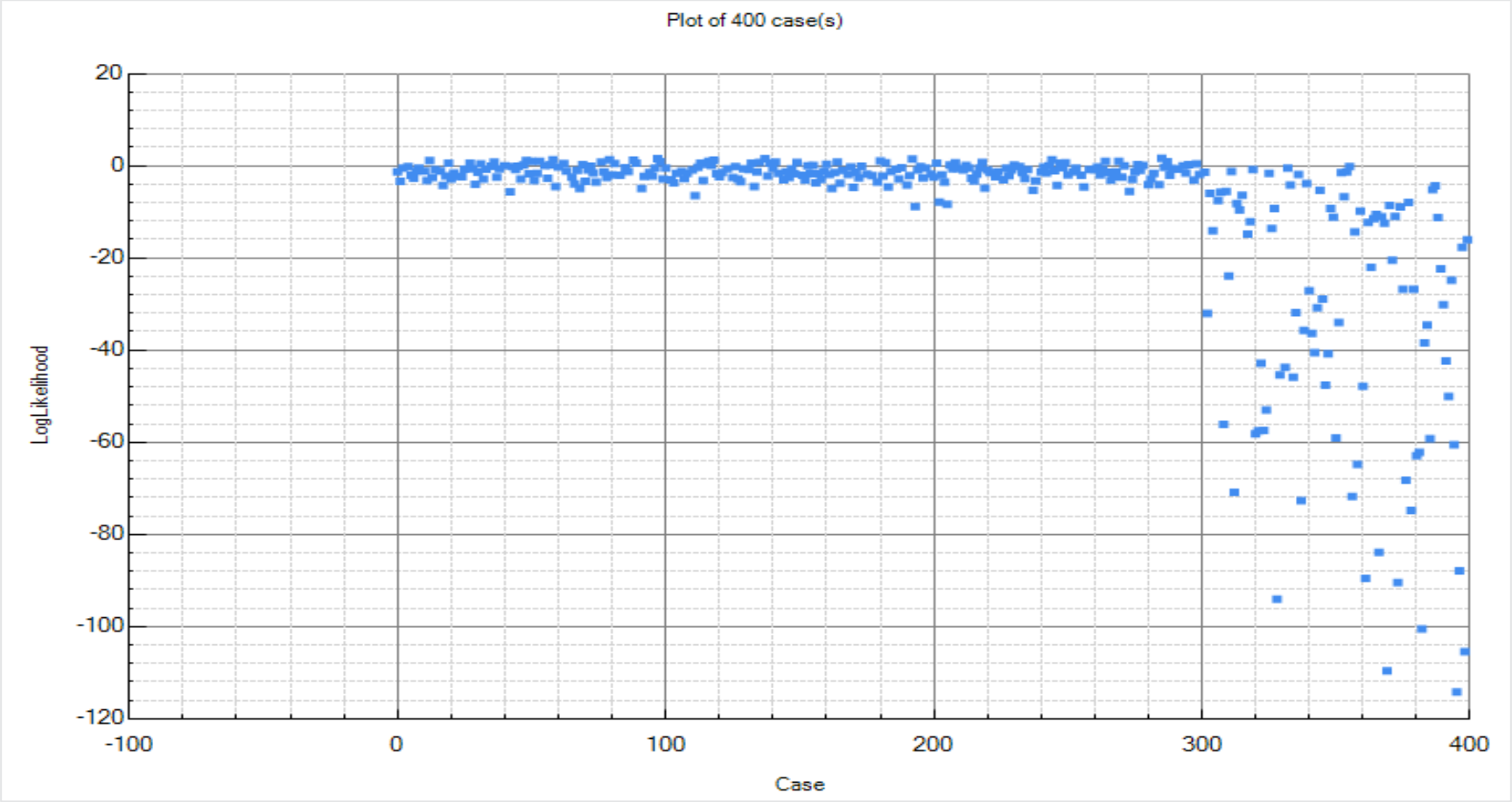






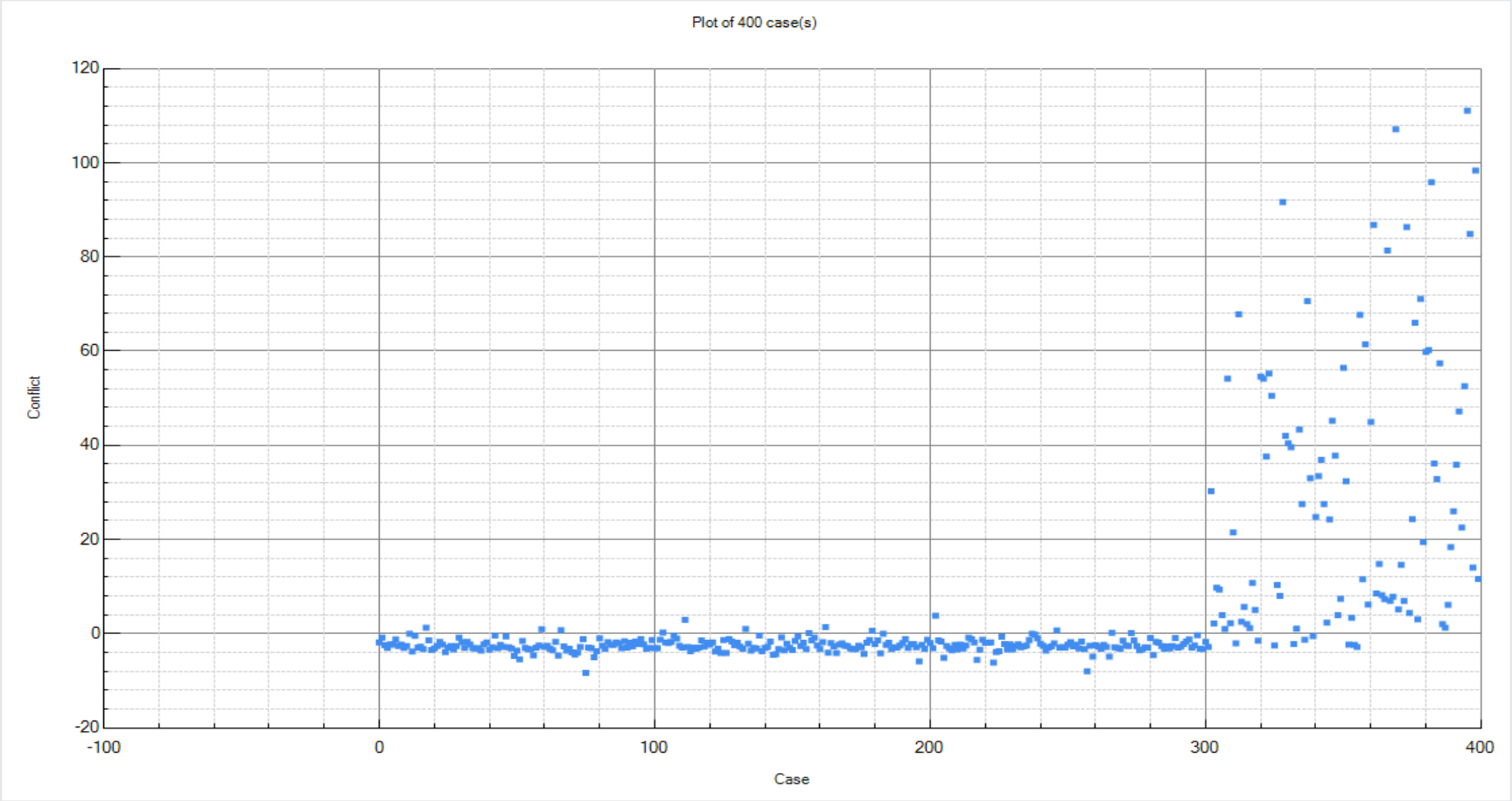
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Multivariate prediction (log-likelihood)



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

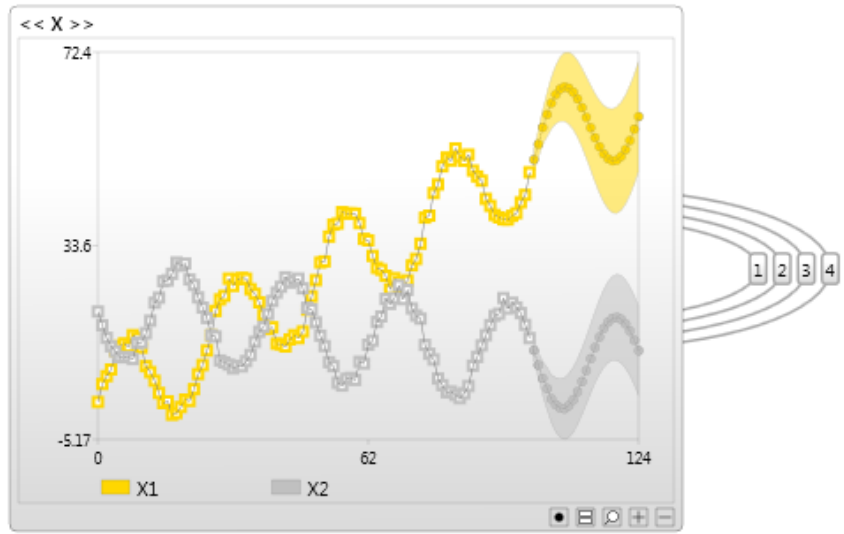
Multivariate prediction (conflict)



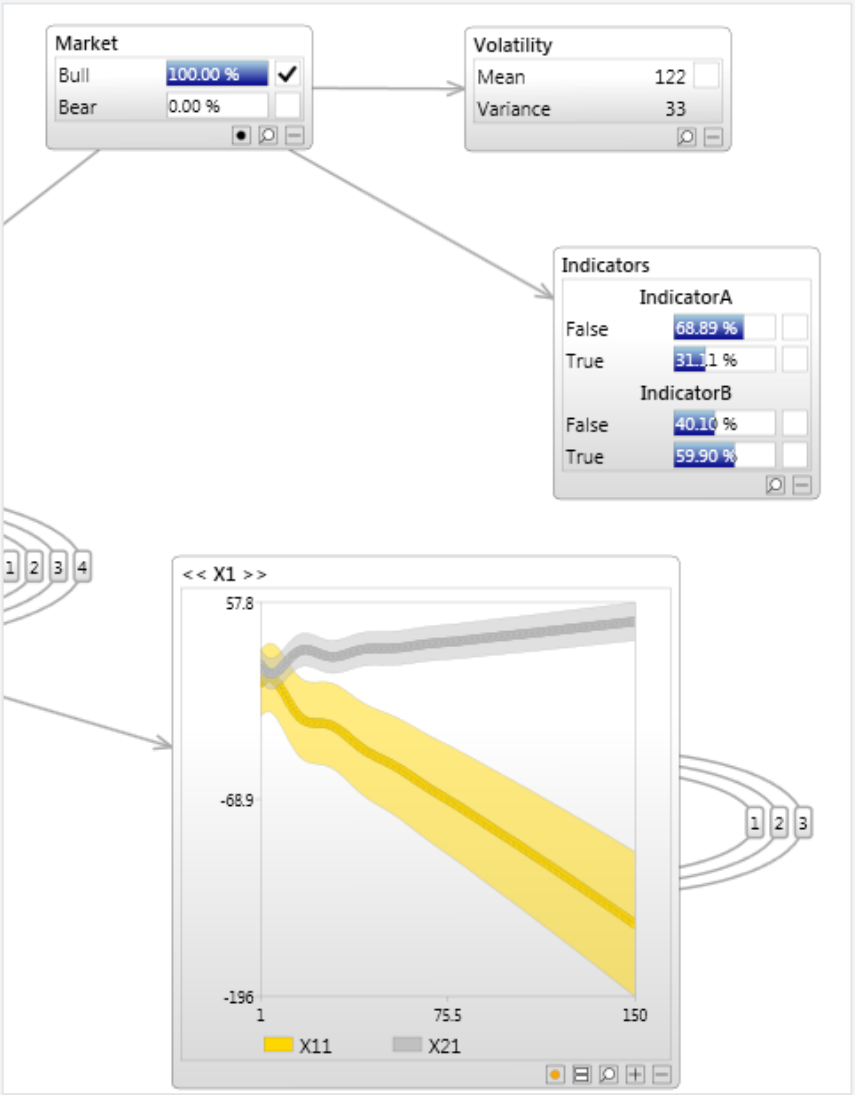
Website: www.BayesServer.com
Email: john.sandiford@BayesServer.com
Twitter: @BayesServer

Time series

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



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



Questions