

# 1. BayesNet

NAME

weka.classifiers.bayes.BayesNet

SYNOPSIS

Bayes Network learning using various search algorithms and quality measures.

Base class for a Bayes Network classifier. Provides datastructures (network structure, conditional probability distributions, etc.) and facilities common to Bayes Network learning algorithms like K2 and B.

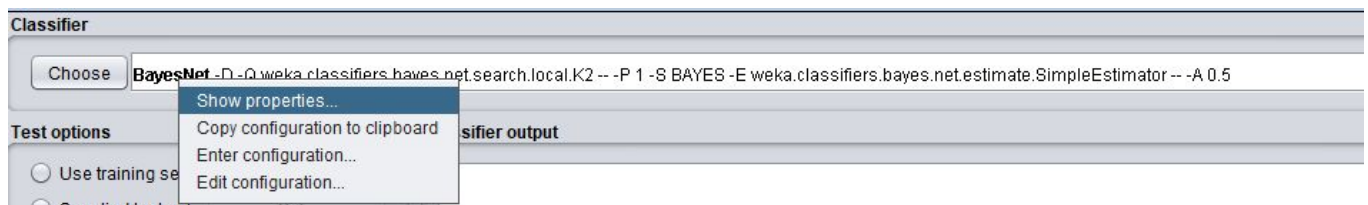
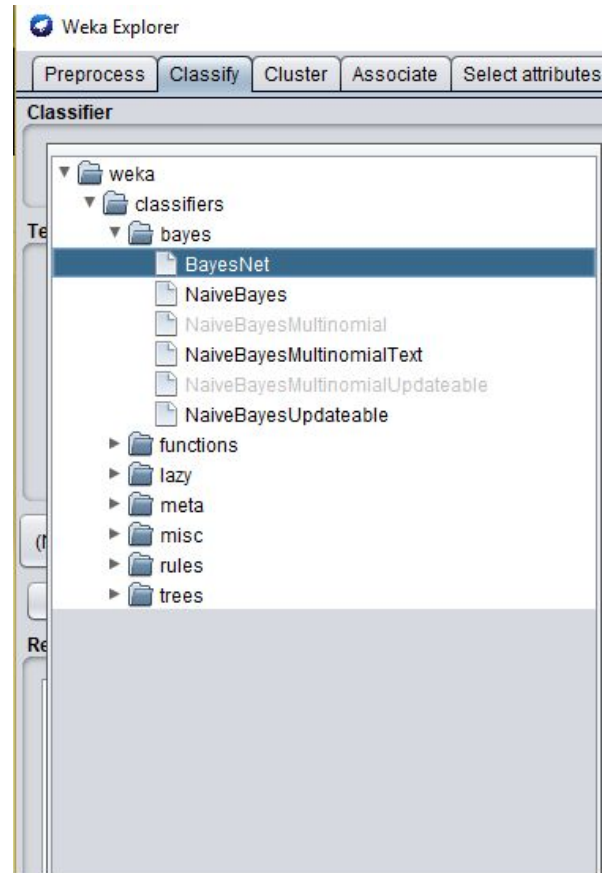
For more information see:

<http://www.cs.waikato.ac.nz/~remco/weka.pdf>

OPTIONS

numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.

batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives



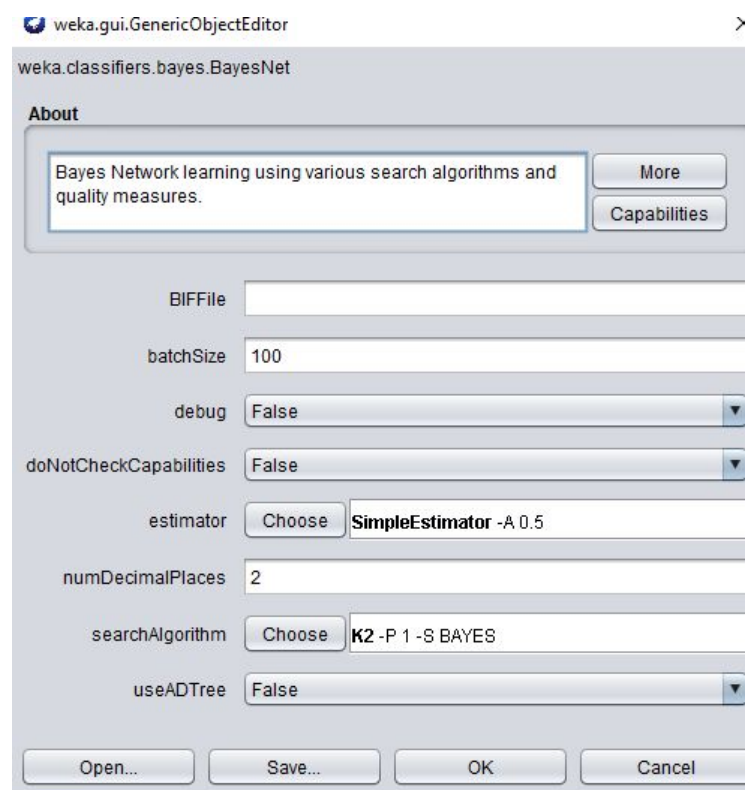
implementations a chance to specify a preferred batch size.

estimator -- Select Estimator algorithm for finding the conditional probability tables of the Bayes Network.

debug -- If set to true, classifier may output additional info to the console.

searchAlgorithm -- Select method used for searching network structures.

doNotCheckCapabilities -- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime).



BIFFile -- Set the name of a file in BIF XML format. A Bayes network learned from data can be compared with the Bayes network represented by the BIF file. Statistics calculated are o.a. the number of missing and extra arcs.

useADTree -- When ADTree (the data structure for increasing speed on counts, not to be confused with the classifier under the same name) is used learning time goes down typically. However, because ADTrees are memory intensive, memory problems may occur. Switching this option off makes the structure learning algorithms slower, and run with less memory. By default, ADTrees are used.

Here is example of Bayes result using Breast-Cancer dataset.

**Weka Explorer**

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

**Classifier**

Choose **BayesNet** -D -Q weka.classifiers.bayes.net.search.local.K2 --P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator --A 0.5

**Test options**

☐ Use training set  
☐ Supplied test set Set...  
☒ Cross-validation Folds **10**  
☐ Percentage split % 66  
 More options...

(Nom) Class

Start Stop

**Result list (right-click for options)**

23:56:46 - bayes.BayesNet

**Classifier output**

```

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      206           72.028 %
Incorrectly Classified Instances     80           27.972 %
Kappa statistic                     0.2919
Mean absolute error                  0.3297
Root mean squared error              0.4566
Relative absolute error              78.7898 %
Root relative squared error          99.9047 %
Total Number of Instances           286

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
               0,841    0,565    0,779      0,841    0,809      0,295    0,698    0,833    no-
               0,435    0,159    0,536      0,435    0,481      0,295    0,698    0,510    rec
Weighted Avg.   0,720    0,444    0,707      0,720    0,711      0,295    0,698    0,737

=== Confusion Matrix ===
  a  b  <-- classified as
169 32 | a = no-recurrence-events
 48 37 | b = recurrence-events
  
```

**Status**

Log

## 2. Naive Bayes

NAME

weka.classifiers.bayes.NaiveBayes

### SYNOPSIS

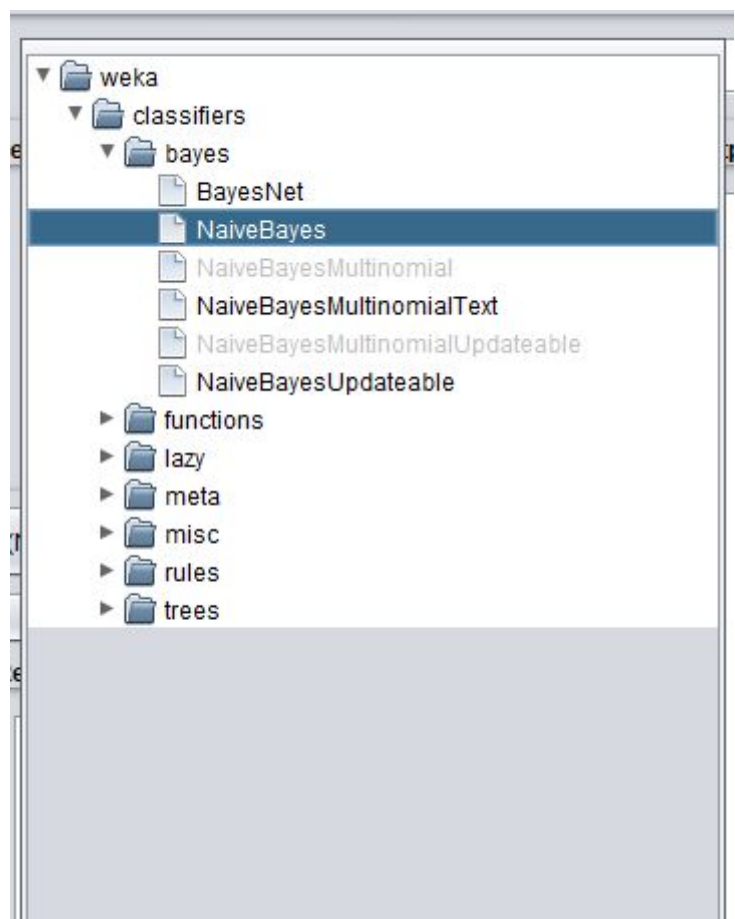
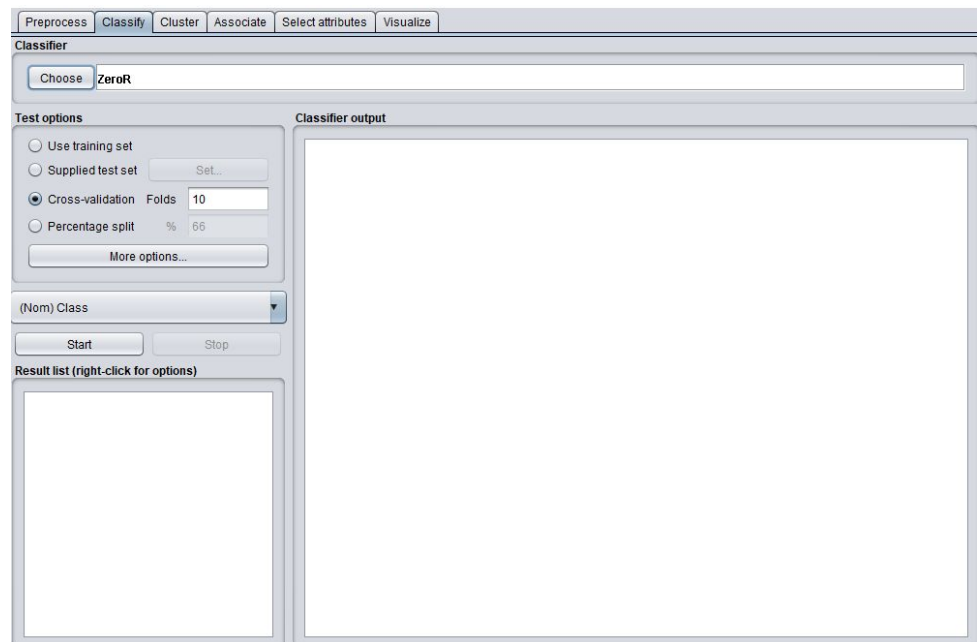
Class for a Naive Bayes classifier using estimator classes.

Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an

UpdateableClassifier (which in typical usage are initialized with zero training instances) -- if you need the UpdateableClassifier functionality, use the NaiveBayesUpdateable classifier. The NaiveBayesUpdateable classifier will use a default precision of 0.1 for numeric attributes when buildClassifier is called with zero training instances.

For more information on Naive Bayes classifiers, see

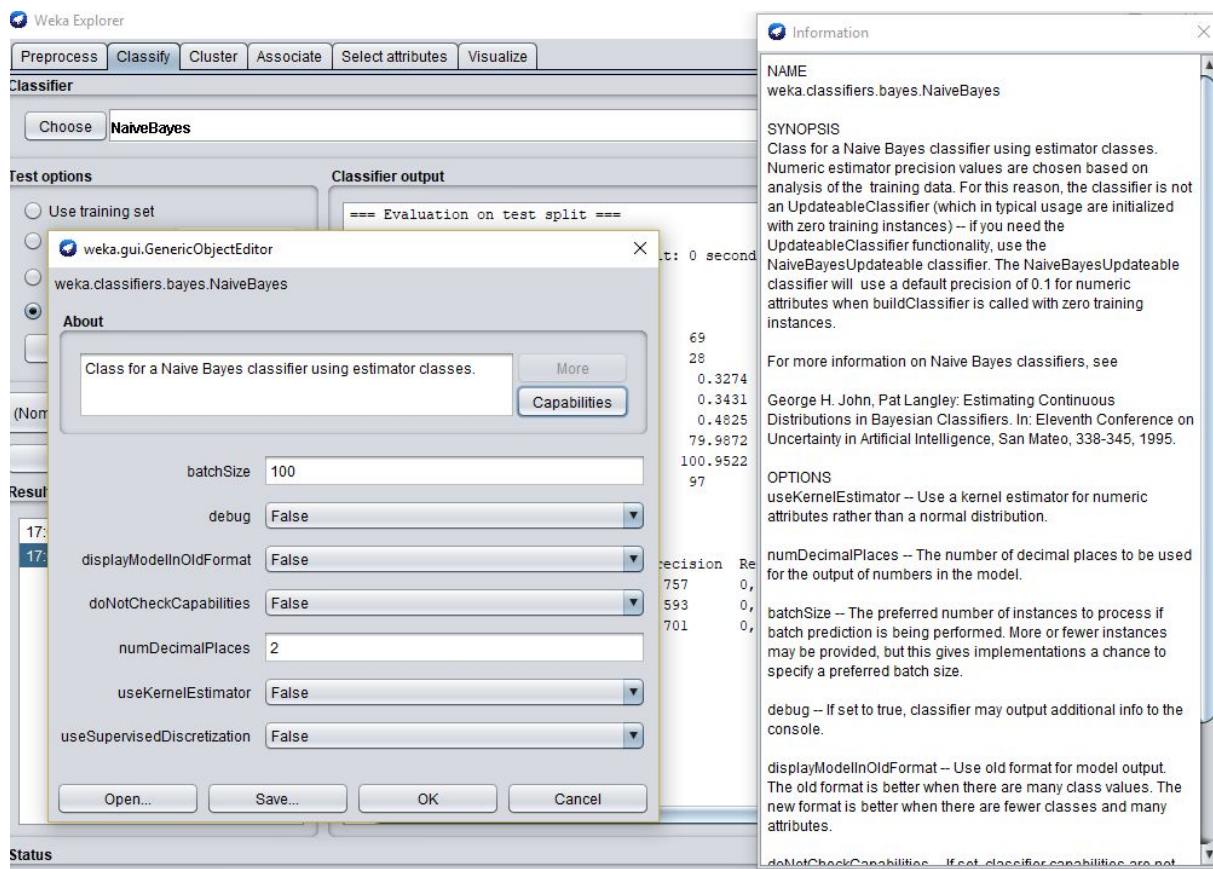
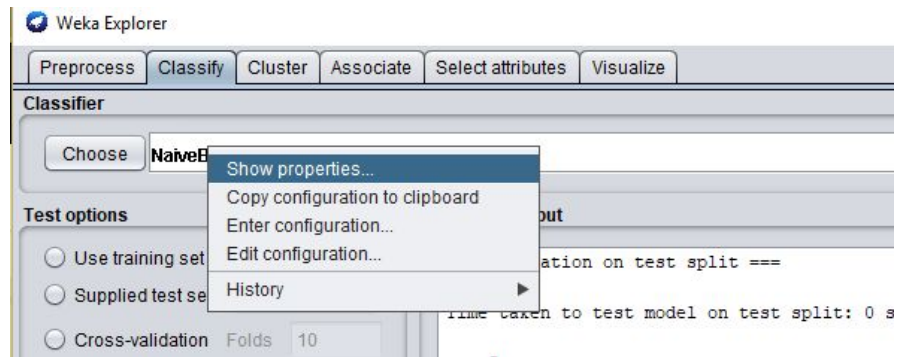
George H. John, Pat Langley:  
Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995.



## OPTIONS

`useKernelEstimator` --  
Use a kernel estimator for numeric attributes rather than a normal distribution.

`numDecimalPlaces` --  
The number of decimal places to be used for the output of numbers in the model.



`batchSize` -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.

`debug` -- If set to true, classifier may output additional info to the console.

`displayModelInOldFormat` -- Use old format for model output. The old format is better when there are many class values. The new format is better when there are fewer classes and many attributes.

doNotCheckCapabilities -- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime).

useSupervisedDiscretization -- Use supervised discretization to convert numeric attributes to nominal ones.

Here is an example of NaiveBayes computing of Breast-Cancer data.

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section on the left shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section on the right displays the results of the stratified cross-validation.

**Test options**

- ☐ Use training set
- ☐ Supplied test set (Set...)
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66
- More options...

**Classifier output**

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      205          71.6783 %
Incorrectly Classified Instances    81           28.3217 %
Kappa statistic                    0.2857
Mean absolute error                 0.3272
Root mean squared error             0.4534
Relative absolute error             78.2086 %
Root relative squared error         99.1872 %
Total Number of Instances          286

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
          0,836   0,565   0,778     0,836   0,806     0,288    0,701    0,837    no-
          0,435   0,164   0,529     0,435   0,477     0,288    0,701    0,514    rec
Weighted Avg.   0,717   0,446   0,704     0,717   0,708     0,288    0,701    0,741

=== Confusion Matrix ===

  a  b  <-- classified as
168 33 |  a = no-recurrence-events
 48 37 |  b = recurrence-events
```

**Result list (right-click for options)**

- 17:05:35 - bayes.NaiveBayes

**Status**