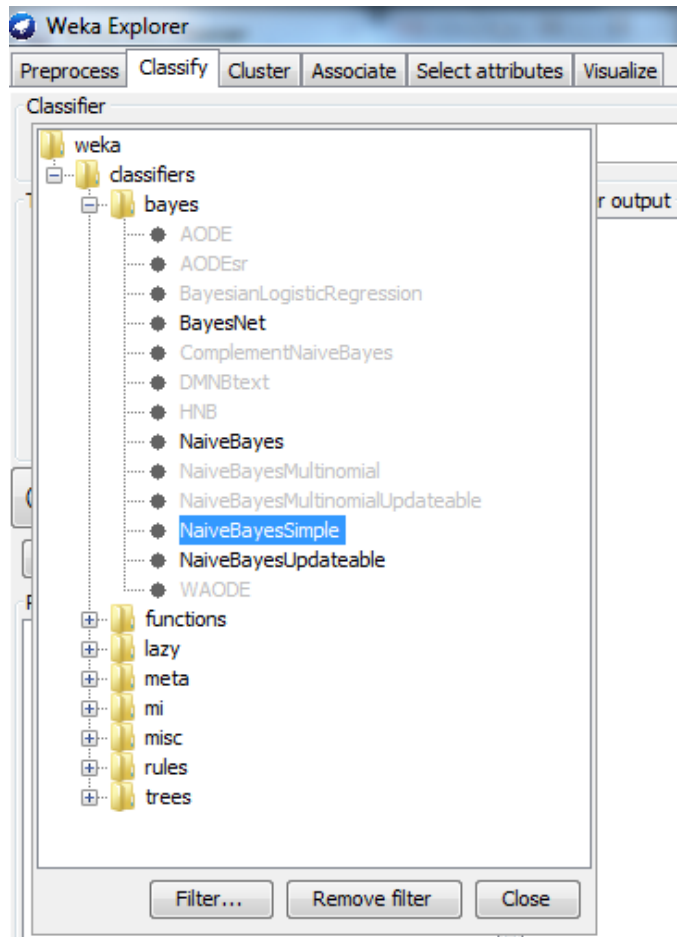




# VARIATIONS IN NAIVE BAYES IMPLEMENTATIONS

**SYRACUSE UNIVERSITY**  
School of Information Studies

# WEKA'S IMPLEMENTATION OF NAIVE BAYES ALGORITHM



Many different versions of naive Bayes' algorithms.

NaiveBayesSimple

NaiveBayes

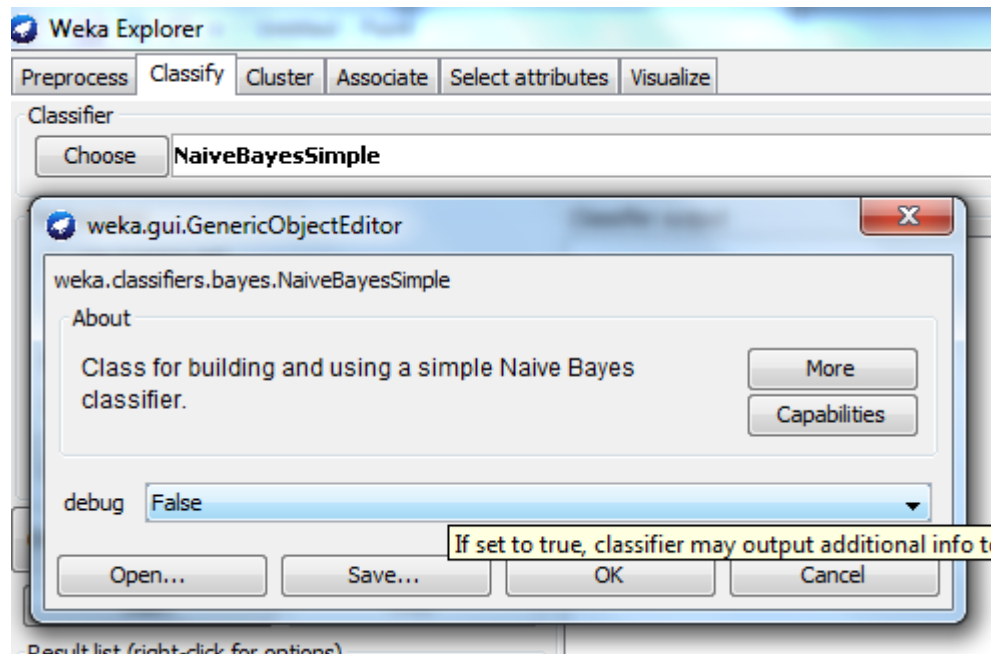
NaiveBayesUpdateable

NaiveBayesMultinomial

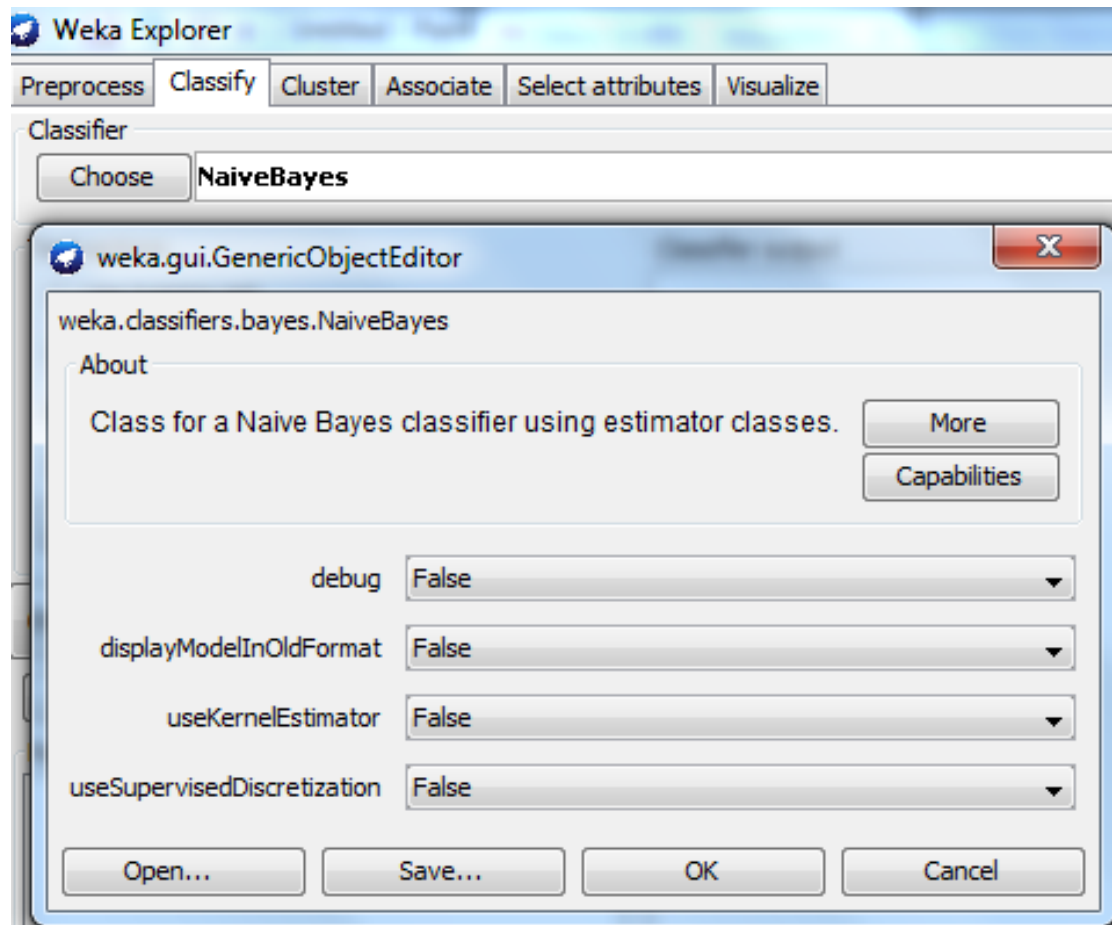
What are their differences?

# NAIVE BAYES SIMPLE

**NaiveBayesSimple** assumes all numeric variables follow normal distribution, and thus does not allow discretization. There are no performance-related parameters to tune in this algorithm.



# NAIVE BAYES



# NAIVE BAYES

**NaiveBayes** allows you to choose from three different methods to handle numeric variables.

Method 1: Use default parameter setting; assume they follow normal distribution (same as `NaiveBayesSimple`).

Method 2: Turn on the “`useKernelEstimator`” option to use kernel estimation for numeric variables. The kernel method does not assume normal distribution.

Method 3: Turn on the “`useSupervisedDiscretization`” option to discretize numeric variables.

Which method to choose?

If your numeric variables follow normal distribution, choose method 1 or `NaiveBayesSimple`.

Otherwise, try the other two methods.

# NAIVE BAYES UPDATEABLE

Sometimes training data may trickle in, instead of coming in a batch. It is computationally expensive to retrain the entire model whenever a new training example is added to the training set.

**NaiveBayesUpdateable** can take new training examples and “update” the existing model without complete retraining.

# NAIVE BAYES MULTINOMIAL

This algorithm is particularly designed for text categorization. For implementation details, see the supplementary reading [mitchell-nb-text-classifier.pdf](#), an excerpt from Tom Mitchell's textbook on machine learning.

# THE E1071 PACKAGE IN R

Library(e1071)

<https://cran.r-project.org/web/packages/e1071/e1071.pdf>



# BAYES' THEOREM IN THE NEWS

*MIT Technology Review*: “How Statisticians Found Air France Flight 447 Two Years After It Crashed Into Atlantic”

<http://www.technologyreview.com/view/527506/how-statisticians-found-air-france-flight-447-two-years-after-it-crashed-into-atlantic/>

NPR News: “Can a 250-Year-Old Mathematical Theorem Find a Missing Plane?”

<http://www.npr.org/blogs/thetwo-way/2014/03/25/294390476/can-a-250-year-old-mathematical-theorem-find-a-missing-plane>