

# Machine Learning with WEKA

Eibe Frank

Department of Computer Science,  
University of Waikato, New Zealand

- WEKA: A Machine Learning Toolkit
- The Explorer
  - Classification and Regression
  - Clustering
  - Association Rules
  - Attribute Selection
  - Data Visualization
- The Experimenter
- The Knowledge Flow GUI
- Conclusions

# WEKA: the bird



*Copyright: Martin Kramer (mkramer@wxs.nl)*

# WEKA: the software

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
- Used for research, education, and applications
- Complements “Data Mining” by Witten & Frank
- Main features:
  - ◆ Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods
  - ◆ Graphical user interfaces (incl. data visualization)
  - ◆ Environment for comparing learning algorithms

# WEKA: versions

- There are several versions of WEKA:
  - ◆ WEKA 3.0: “book version” compatible with description in data mining book
  - ◆ WEKA 3.2: “GUI version” adds graphical user interfaces (book version is command-line only)
  - ◆ WEKA 3.3: “development version” with lots of improvements
- This talk is based on the latest snapshot of WEKA 3.3 (soon to be WEKA 3.4)

# WEKA only deals with “flat” files

@relation heart-disease-simplified

@attribute age numeric

@attribute sex { female, male}

@attribute chest\_pain\_type { typ\_angina, asympt, non\_anginal, atyp\_angina}

@attribute cholesterol numeric

@attribute exercise\_induced\_angina { no, yes}

@attribute class { present, not\_present}

@data

63,male,typ\_angina,233,no,not\_present

67,male,asympt,286,yes,present

67,male,asympt,229,yes,present

38,female,non\_anginal,?,no,not\_present

...



Flat file in  
ARFF format

# WEKA only deals with “flat” files

@relation heart-disease-simplified

@attribute age numeric

@attribute sex { female, male}

@attribute chest\_pain\_type { typ\_angina, asympt, non\_anginal, atyp\_angina}

@attribute cholesterol numeric

@attribute exercise\_induced\_angina { no, yes}

@attribute class { present, not\_present}

@data

63,male,typ\_angina,233,no,not\_present

67,male,asympt,286,yes,present

67,male,asympt,229,yes,present

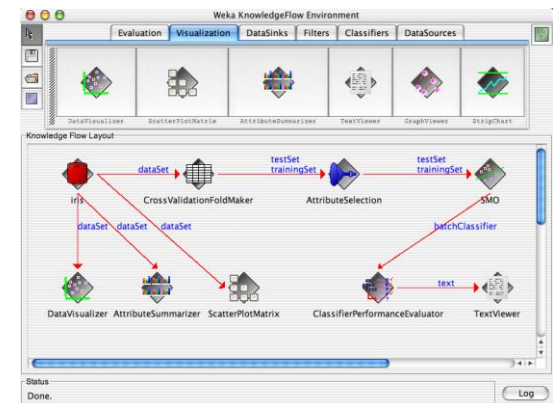
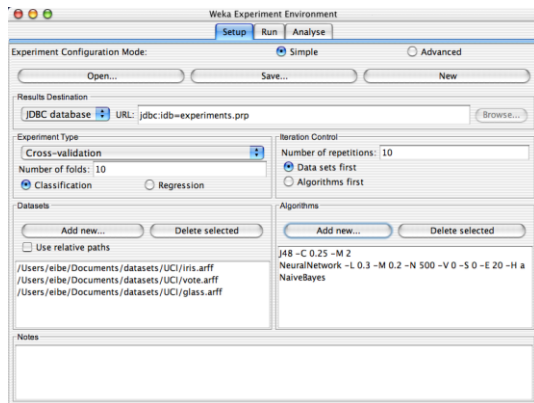
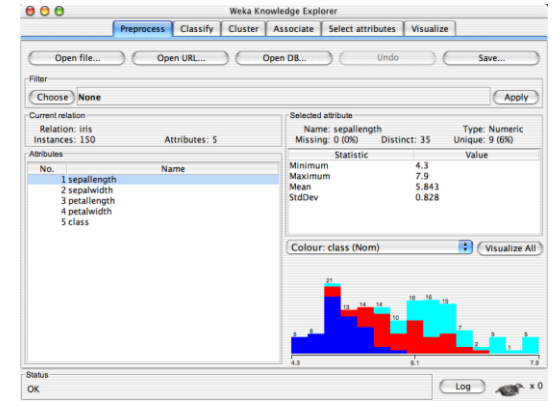
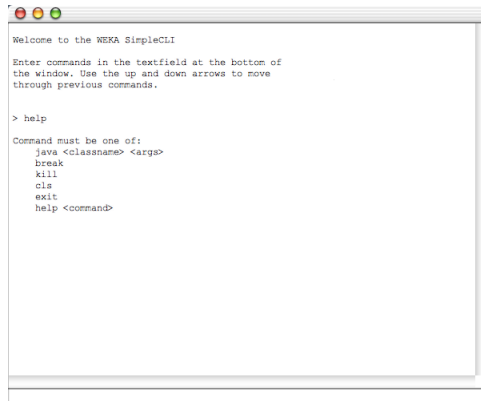
38,female,non\_anginal,?,no,not\_present

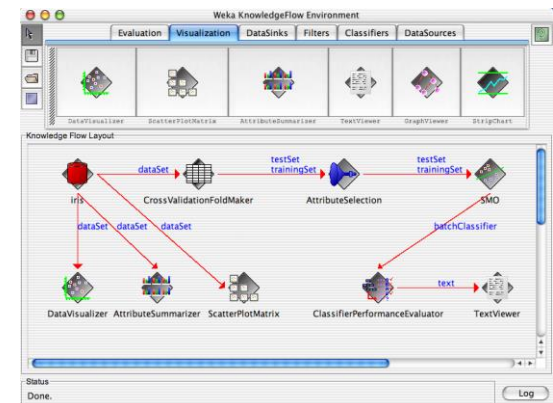
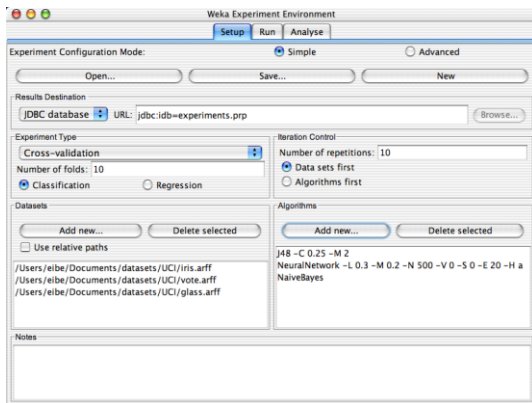
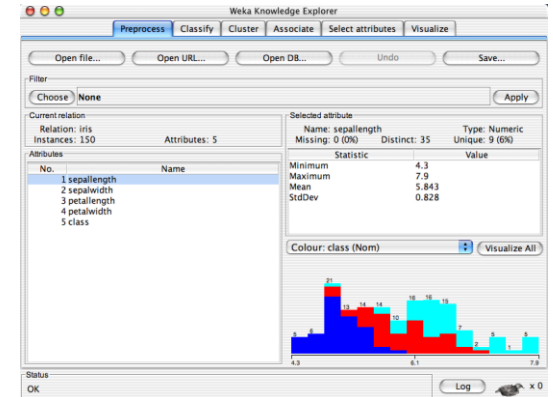
...

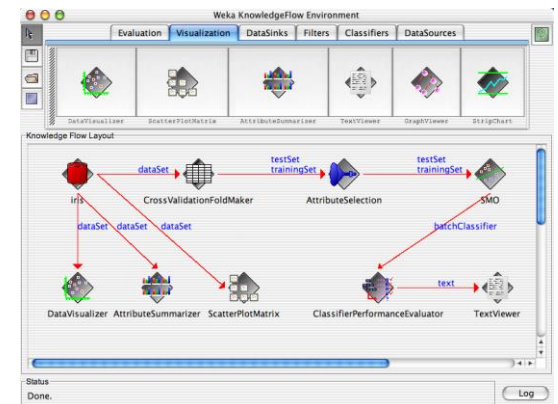
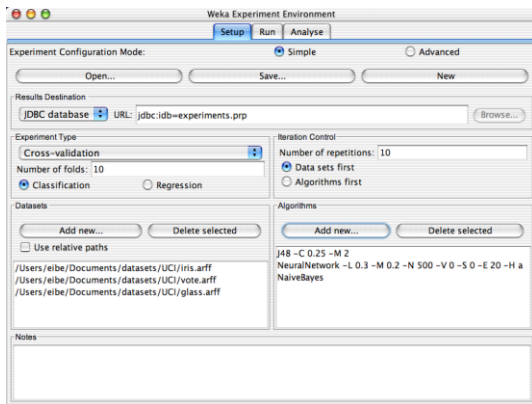
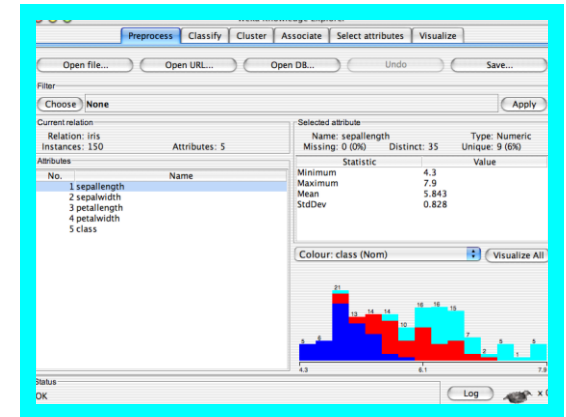
numeric attribute

nominal attribute









# Explorer: pre-processing the data

- Data can be imported from a file in various formats: ARFF, CSV, C4.5, binary
- Data can also be read from a URL or from an SQL database (using JDBC)
- Pre-processing tools in WEKA are called “filters”
- WEKA contains filters for:
  - ◆ Discretization, normalization, resampling, attribute selection, transforming and combining attributes, ...



# Weka Knowledge Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

None

Apply

Current relation

Relation: None

Instances: None

Attributes: None

Selected attribute

Name: None

Missing: None

Type: None

Distinct: None

Unique: None

Attributes

Empty list for attributes

Empty list for selected attributes



Visualize All

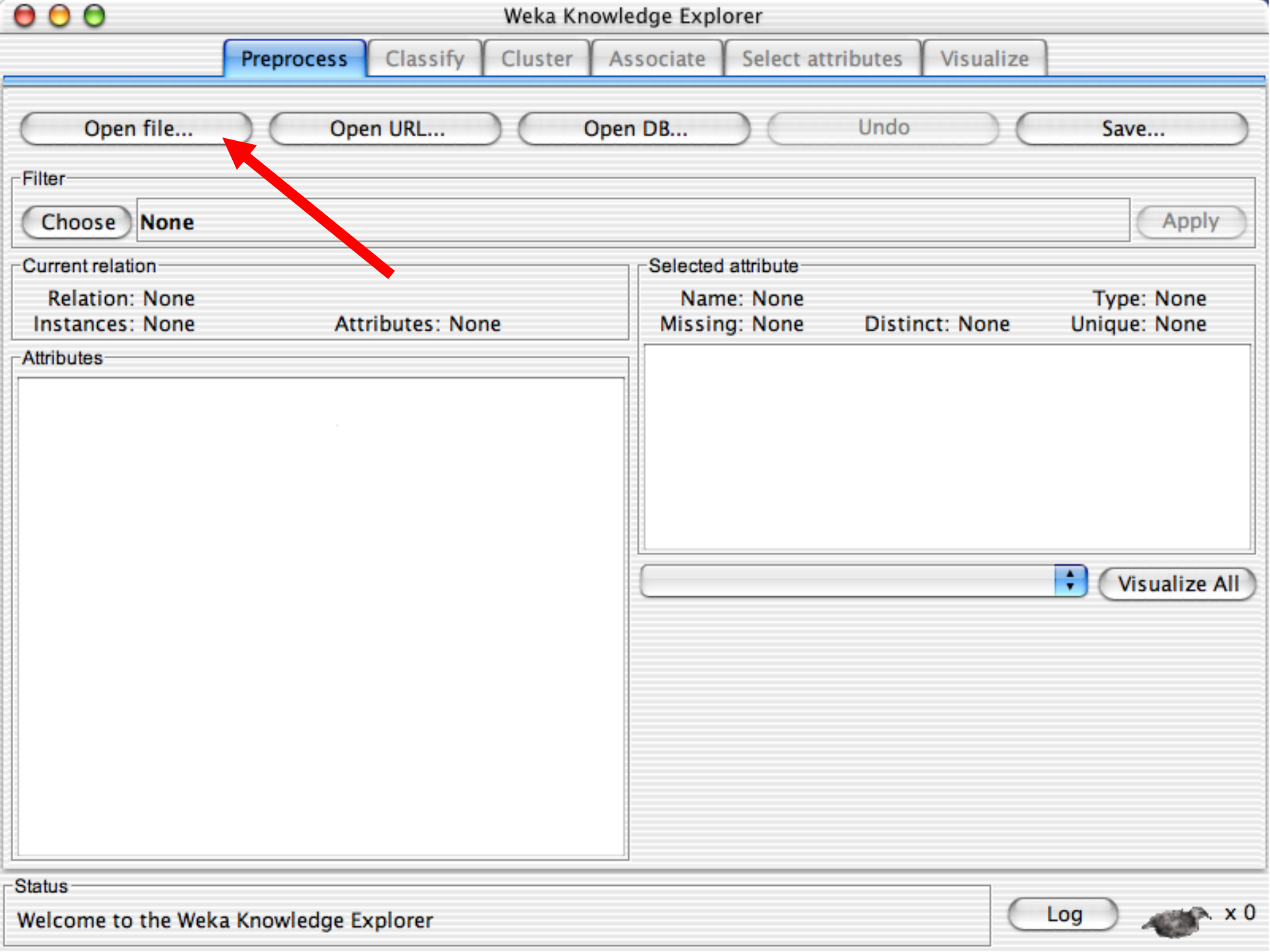
Status

Welcome to the Weka Knowledge Explorer

Log



x 0



# Weka Knowledge Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

None

Apply

Current relation

Relation: None

Instances: None

Attributes: None

Selected attribute

Name: None

Missing: None

Distinct: None

Type: None

Unique: None

Attributes

Visualize All

Status

Welcome to the Weka Knowledge Explorer

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

None

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

| No. | Name        |
|-----|-------------|
| 1   | sepallength |
| 2   | sepalwidth  |
| 3   | petallength |
| 4   | petalwidth  |
| 5   | class       |

Selected attribute

Name: sepallength

Type: Numeric

Missing: 0 (0%)

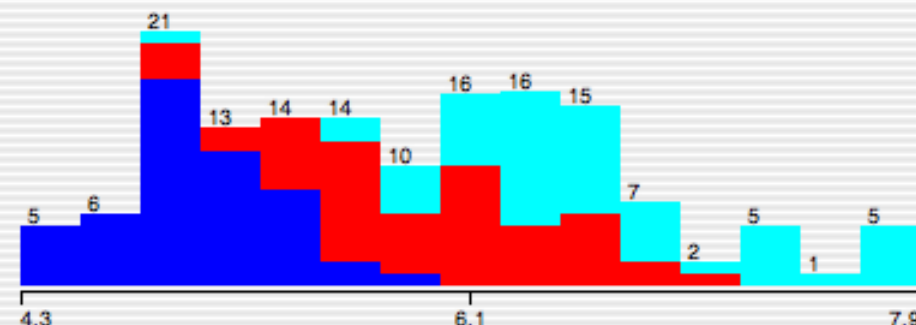
Distinct: 35

Unique: 9 (6%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 4.3   |
| Maximum   | 7.9   |
| Mean      | 5.843 |
| StdDev    | 0.828 |

Colour: class (Nom)

Visualize All



4.3

6.1

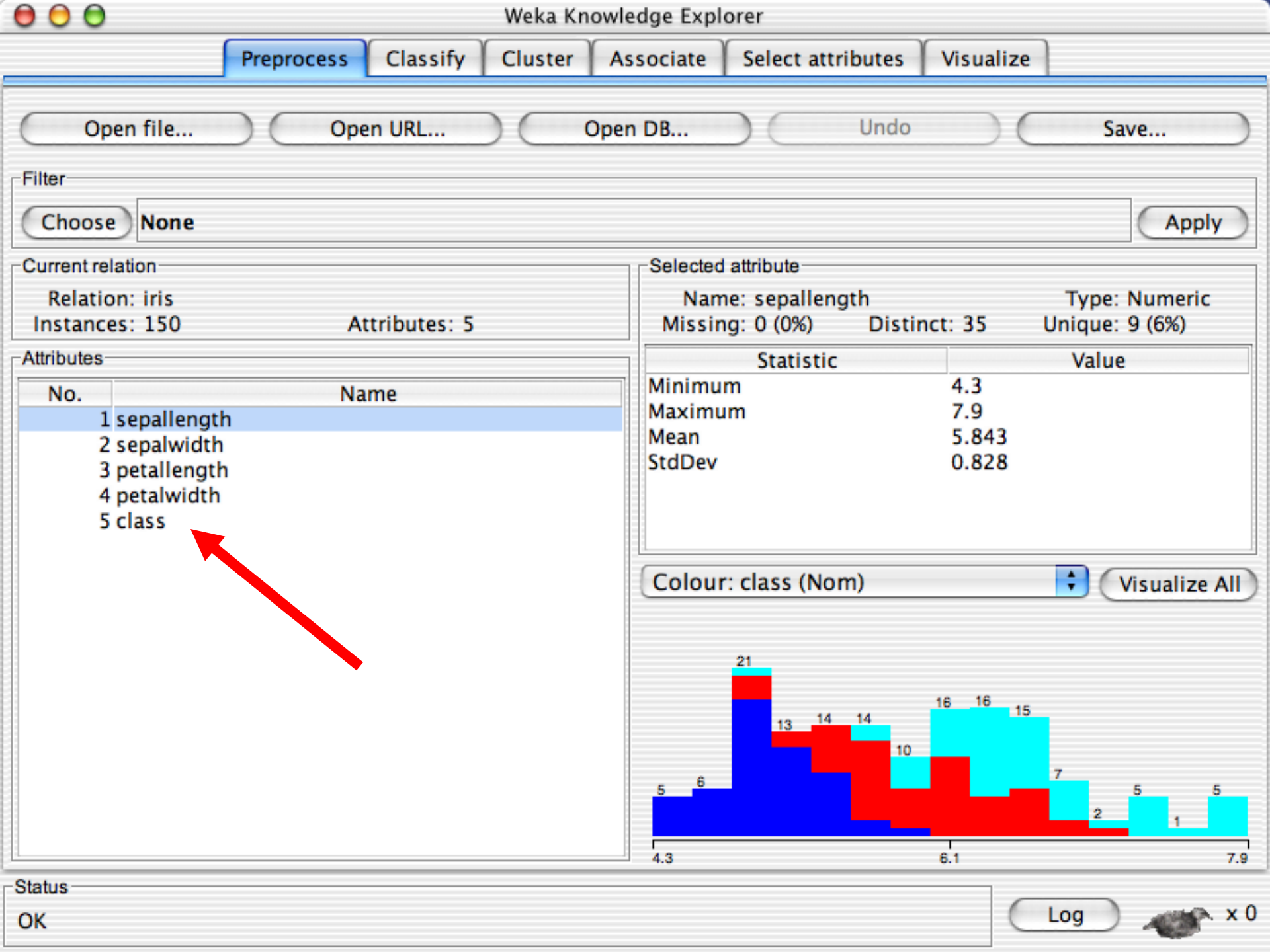
7.9

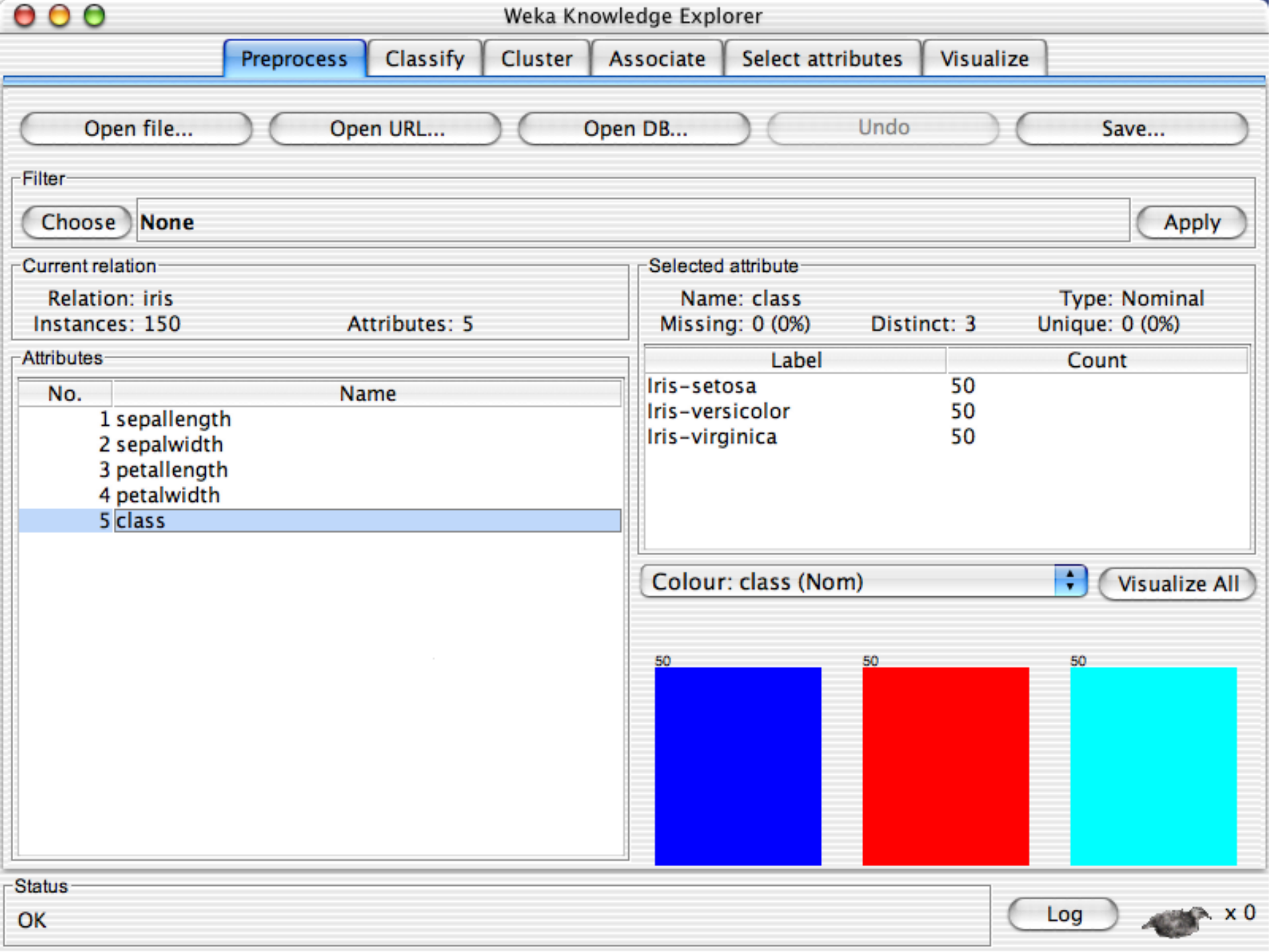
Status

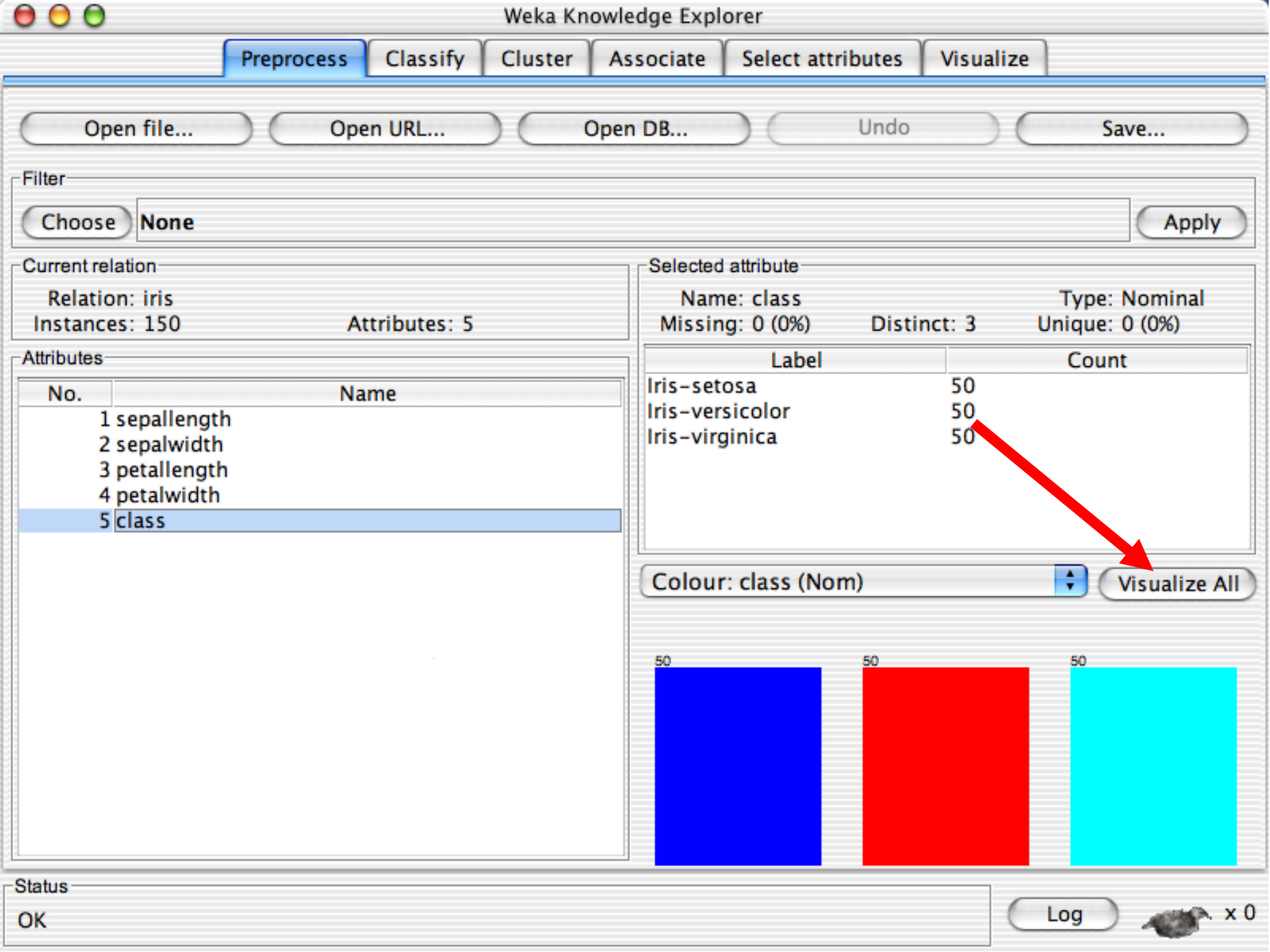
OK

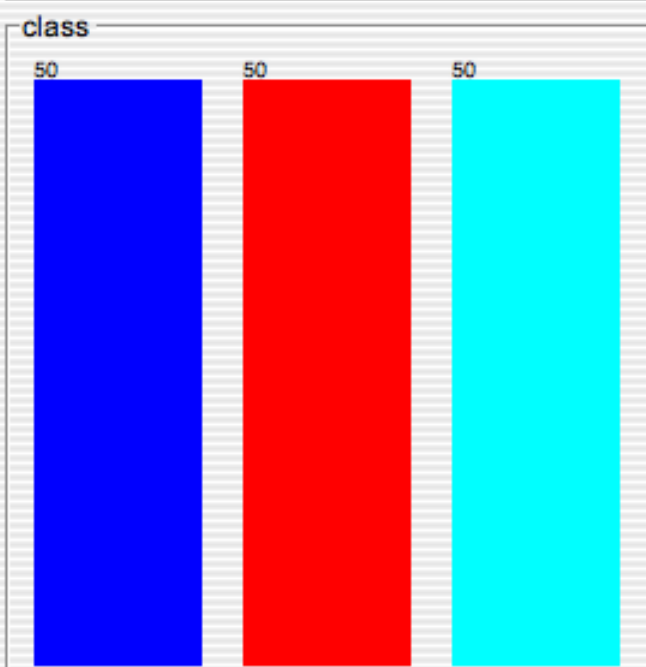
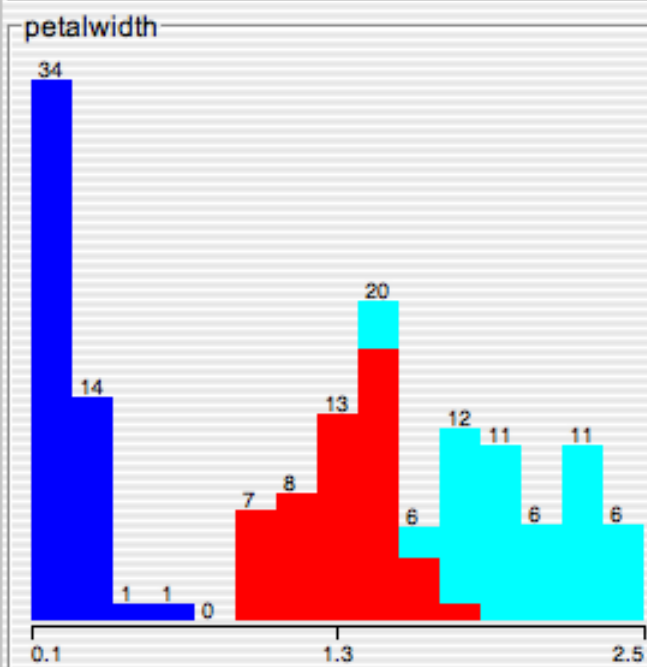
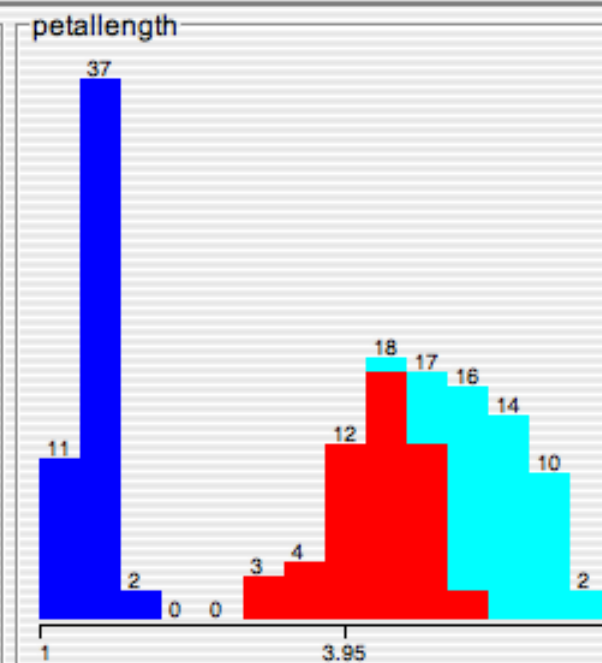
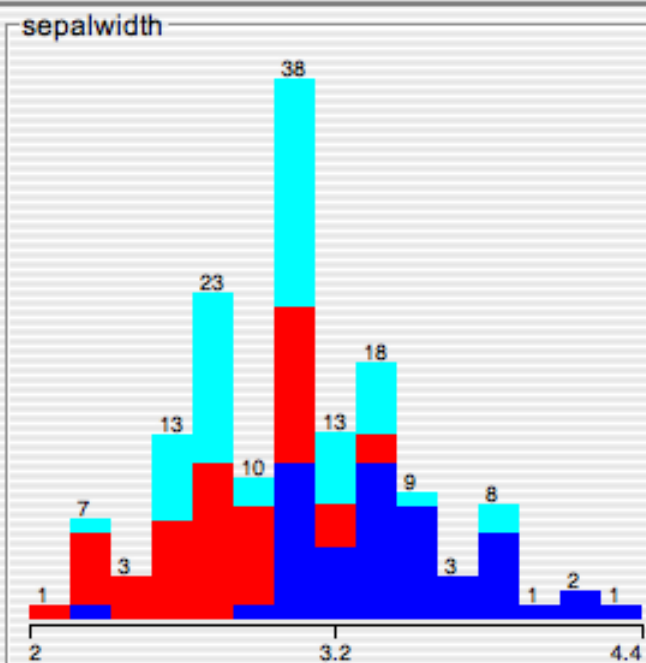
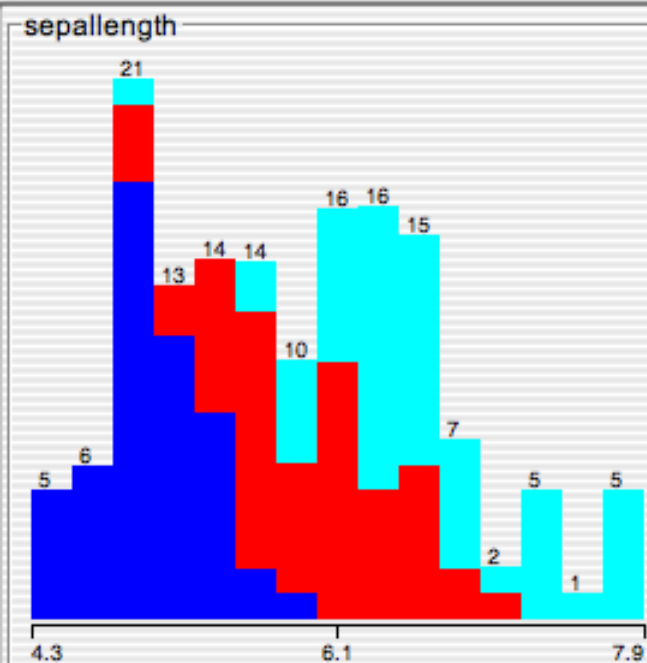
Log

 x 0











Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

None

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

Selected attribute

Name: petal.length

Type: Numeric

Missing: 0 (0%)

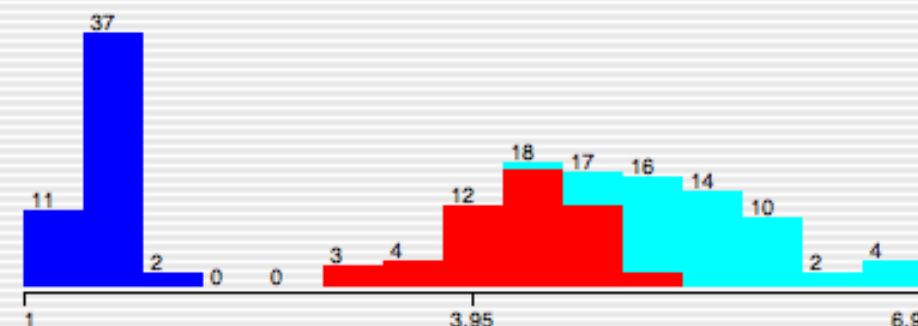
Distinct: 43

Unique: 10 (7%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 1     |
| Maximum   | 6.9   |
| Mean      | 3.759 |
| StdDev    | 1.764 |

Colour: class (Nom)

Visualize All

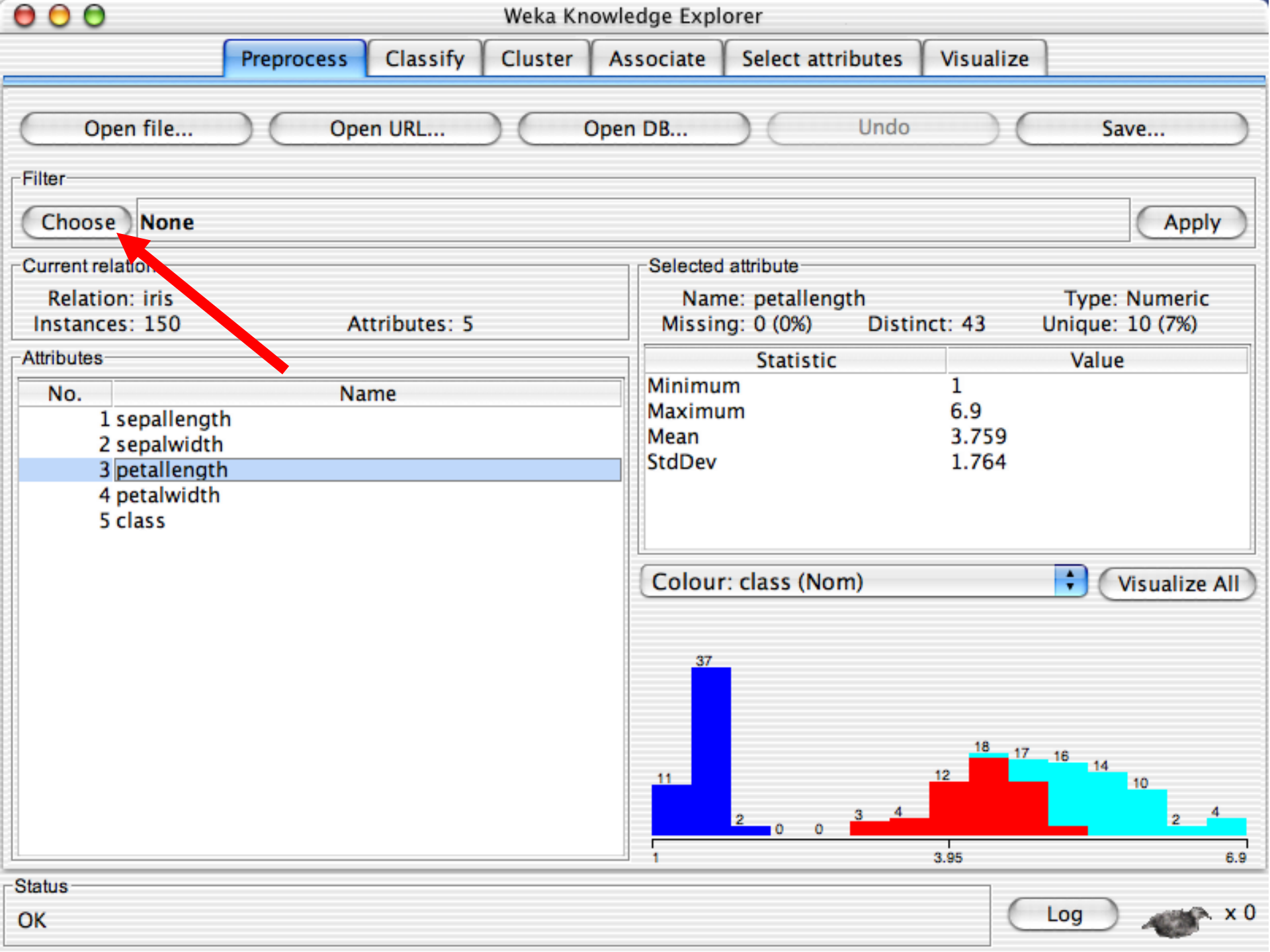


Status

OK

Log

 x 0





Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

weka

- filters
  - unsupervised
    - attribute
    - instance

Apply

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

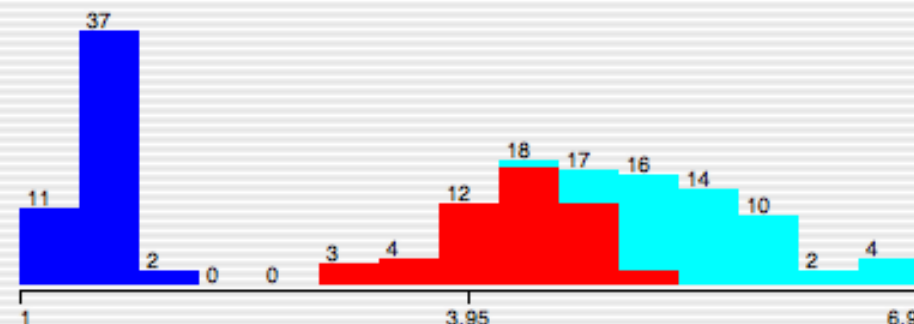
Distinct: 43

Unique: 10 (7%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 1     |
| Maximum   | 6.9   |
| Mean      | 3.759 |
| StdDev    | 1.764 |

Colour: class (Nom)

Visualize All

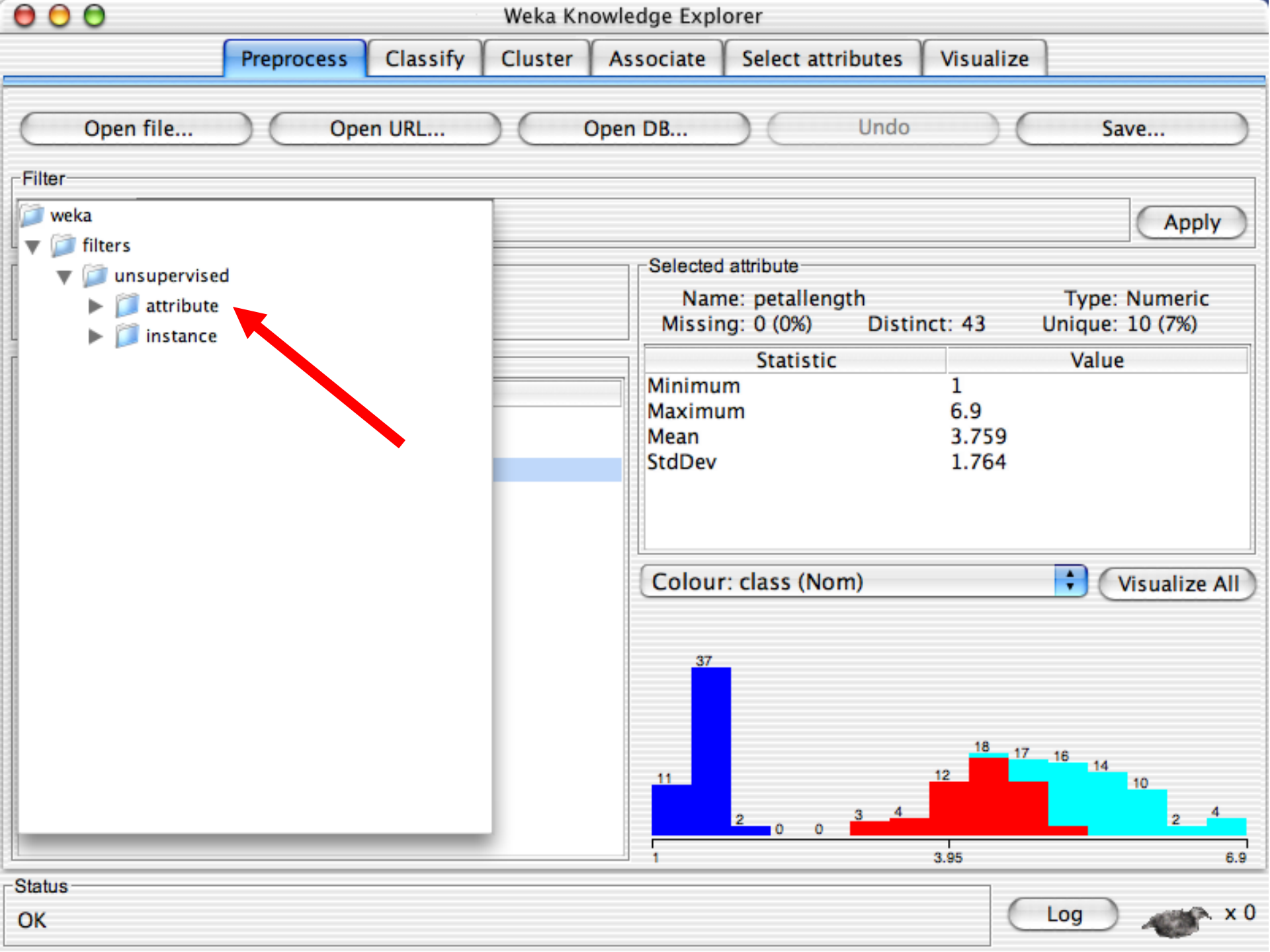


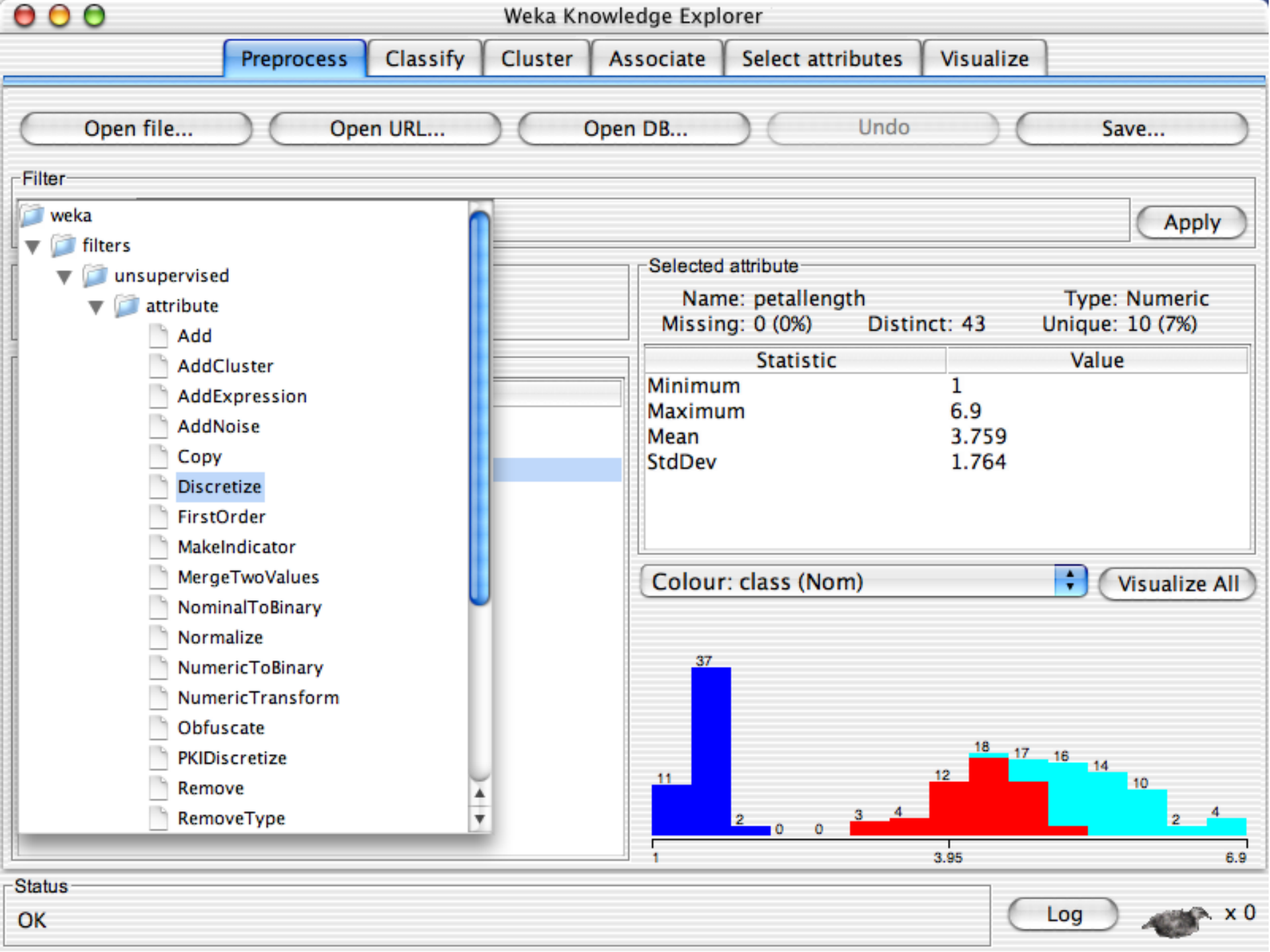
Status

OK

Log

 x 0







Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

Discretize -B 10 -R first-last

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

Selected attribute

Name: petal.length

Type: Numeric

Missing: 0 (0%)

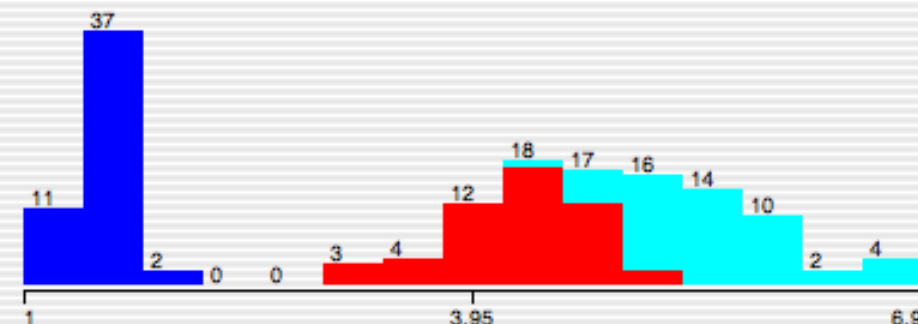
Distinct: 43

Unique: 10 (7%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 1     |
| Maximum   | 6.9   |
| Mean      | 3.759 |
| StdDev    | 1.764 |

Colour: class (Nom)

Visualize All



Status

OK

Log

 x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

Discretize -B 10 -R first-last

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

Selected attribute

Name: petal.length

Type: Numeric

Missing: 0 (0%)

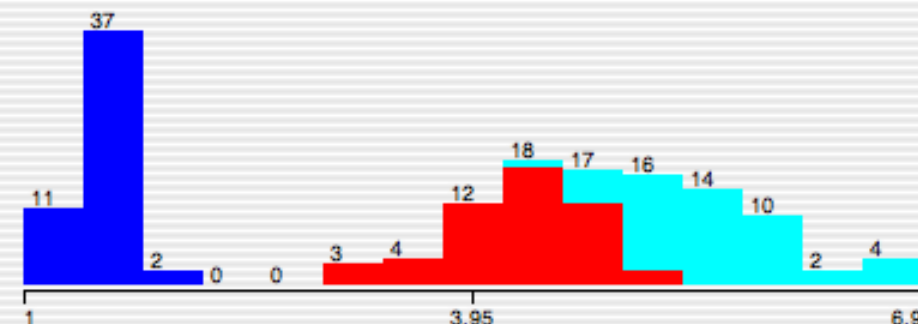
Distinct: 43

Unique: 10 (7%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 1     |
| Maximum   | 6.9   |
| Mean      | 3.759 |
| StdDev    | 1.764 |

Colour: class (Nom)

Visualize All

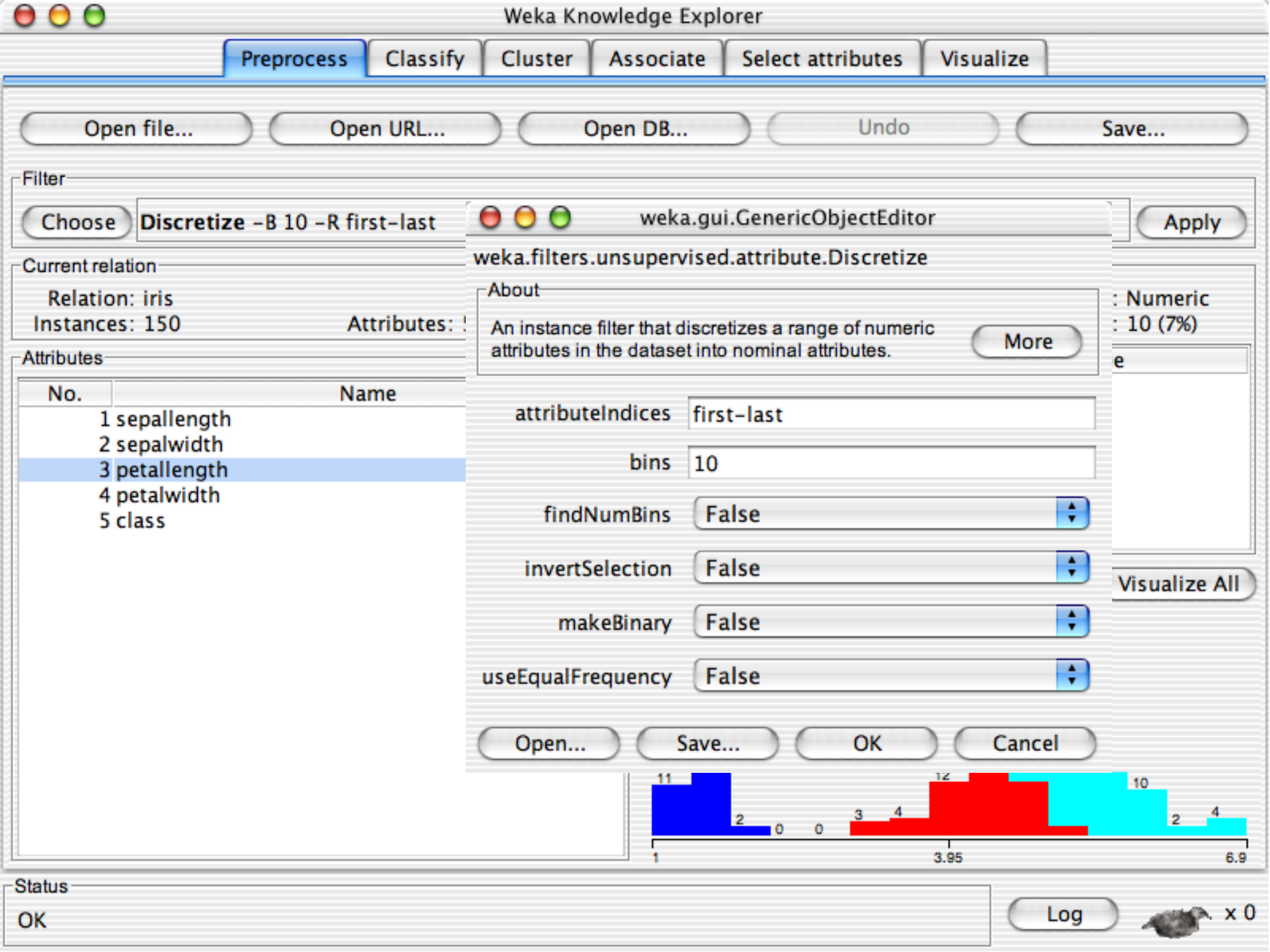


Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

Discretize -B 10 -R first-last

Current relation

Relation: iris

Instances: 150

Attributes:

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About

An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.

More

attributeIndices

first-last

bins

10

findNumBins

False

invertSelection

False

makeBinary

False

useEqualFrequency

False

Open...

Save...

OK

Cancel

Visualize All

11

2

0

0

3

4

14

10

2

4

1

3.95

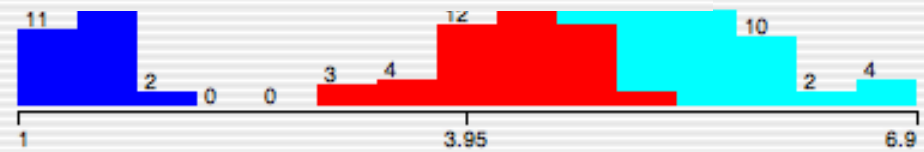
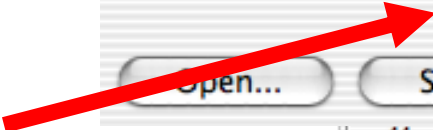
6.9

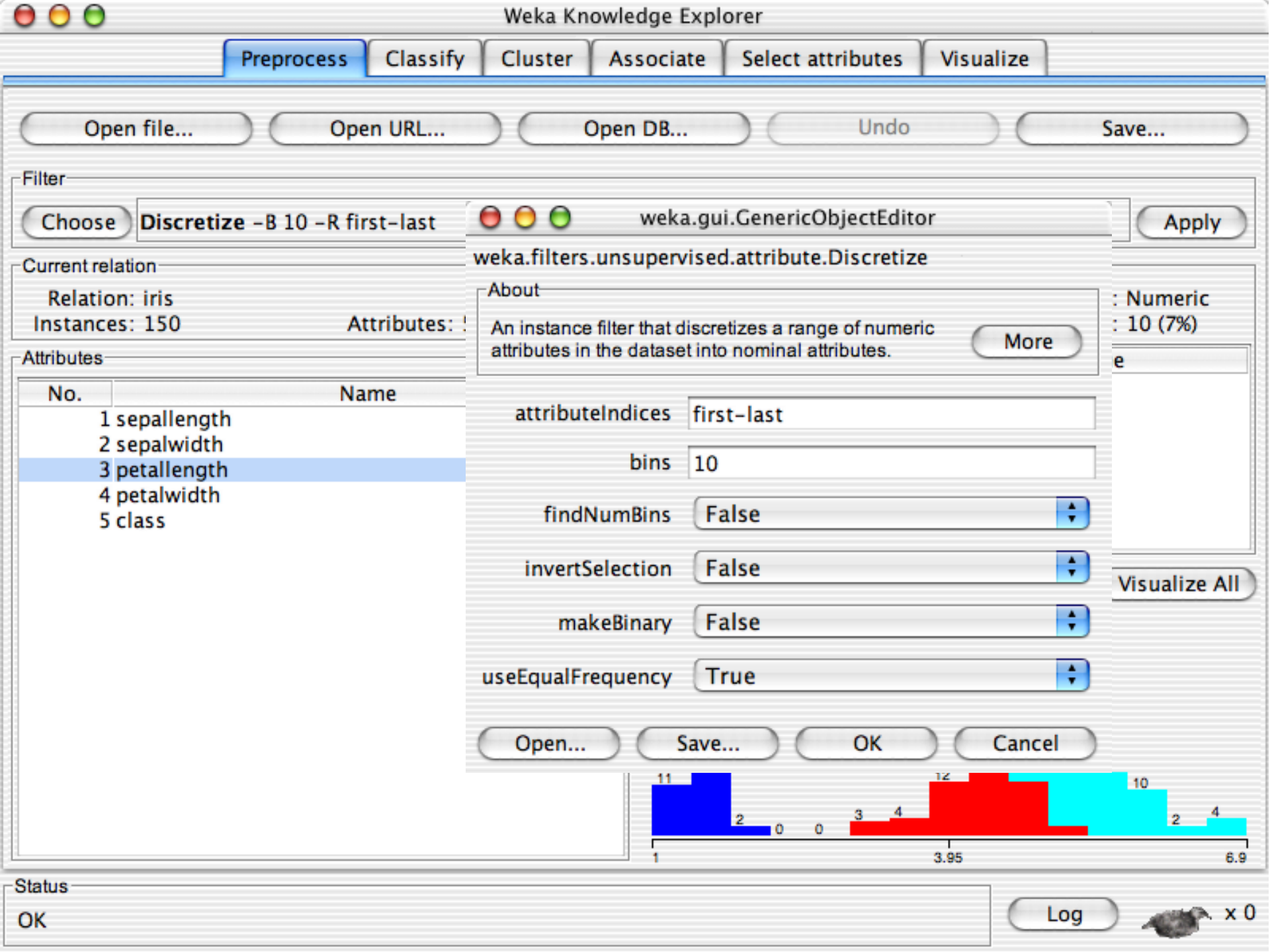
Status

OK

Log

x 0





Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

Discretize -B 10 -R first-last

Current relation

Relation: iris

Instances: 150

Attributes:

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About

An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.

More

attributeIndices

first-last

bins

10

findNumBins

False

invertSelection

False

makeBinary

False

useEqualFrequency

True

Open...

Save...

OK

Cancel

11

2

0

0

3

4

14

10

2

4

1

3.95

6.9

Visualize All

Discretize -B 10 -R first-last

More

Visualize All

Status

OK

Log

x 0



- Preprocess
- Classify
- Cluster
- Associate
- Select attributes
- Visualize

- Open file...
- Open URL...
- Open DB...
- Undo
- Save...

Filter

Choose

Discretize -B 10 -R first-last

Current relation

Relation: iris

Instances: 150

Attributes:

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About

An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.

More

attributeIndices

first-last

bins

10

findNumBins

False

invertSelection

False

makeBinary

False

useEqualFrequency

True

Open...

Save...

OK

Cancel

11

2

0

0

3

4

14

10

2

4

1

3.95

6.9

Visualize All

Discretize -B 10 -R first-last

More

Visualize All



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

Discretize -F -B 10 -R first-last

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

Selected attribute

Name: petal.length

Type: Numeric

Missing: 0 (0%)

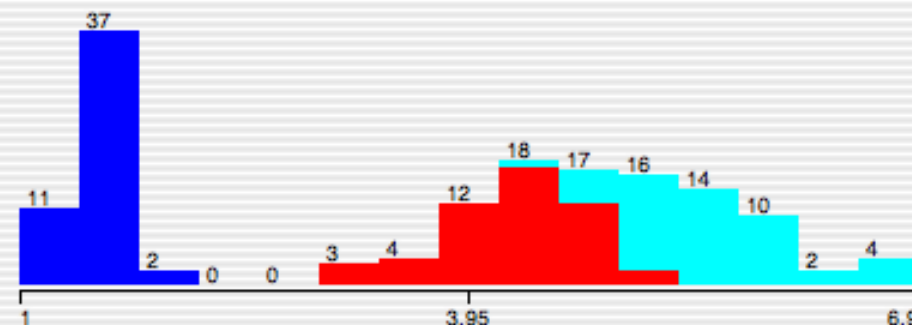
Distinct: 43

Unique: 10 (7%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 1     |
| Maximum   | 6.9   |
| Mean      | 3.759 |
| StdDev    | 1.764 |

Colour: class (Nom)

Visualize All

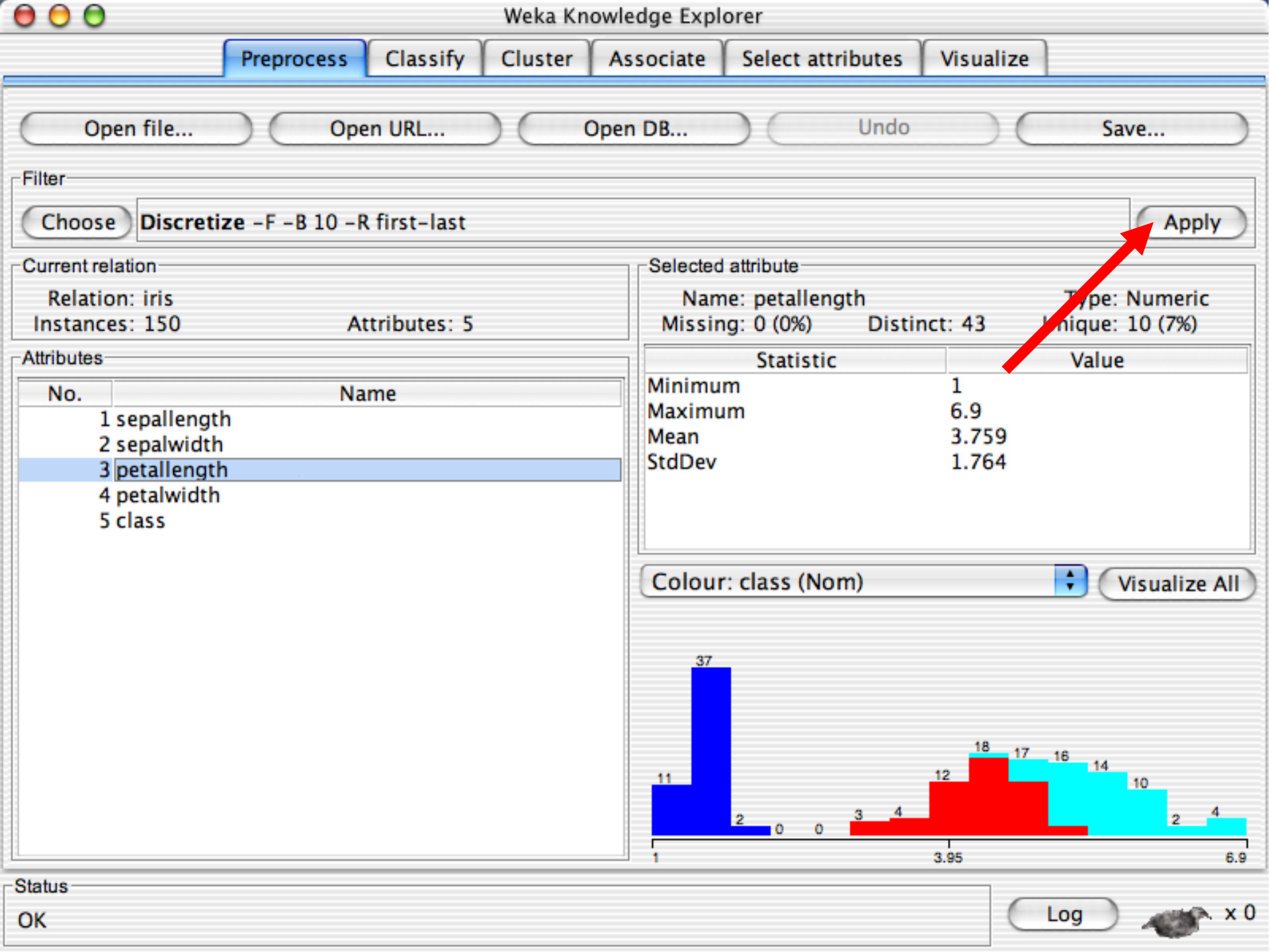


Status

OK

Log







Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

Discretize -F -B 10 -R first-last

Apply

Current relation

Relation: iris-weka.filters.unsupervised.attribute.Disc...

Instances: 150

Attributes: 5

Attributes

| No. | Name         |
|-----|--------------|
| 1   | sepal.length |
| 2   | sepal.width  |
| 3   | petal.length |
| 4   | petal.width  |
| 5   | class        |

Selected attribute

Name: petal.length

Type: Nominal

Missing: 0 (0%)

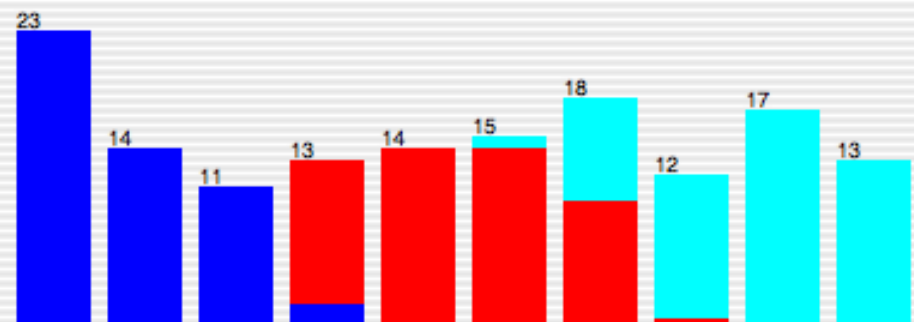
Distinct: 10

Unique: 0 (0%)

| Label         | Count |
|---------------|-------|
| '(-inf-1.45]' | 23    |
| '(1.45-1.55]' | 14    |
| '(1.55-1.8]'  | 11    |
| '(1.8-3.95]'  | 13    |
| '(3.95-4.35]' | 14    |
| '(4.35-4.65]' | 15    |
| '(4.65-5.05]' | 18    |

Colour: class (Nom)

Visualize All



Status

OK

Log

 x 0

# Explorer: building “classifiers”

- Classifiers in WEKA are models for predicting nominal or numeric quantities
- Implemented learning schemes include:
  - ◆ Decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes’ nets, ...
- “Meta”-classifiers include:
  - ◆ Bagging, boosting, stacking, error-correcting output codes, locally weighted learning, ...



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

ZeroR

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)

Classifier output

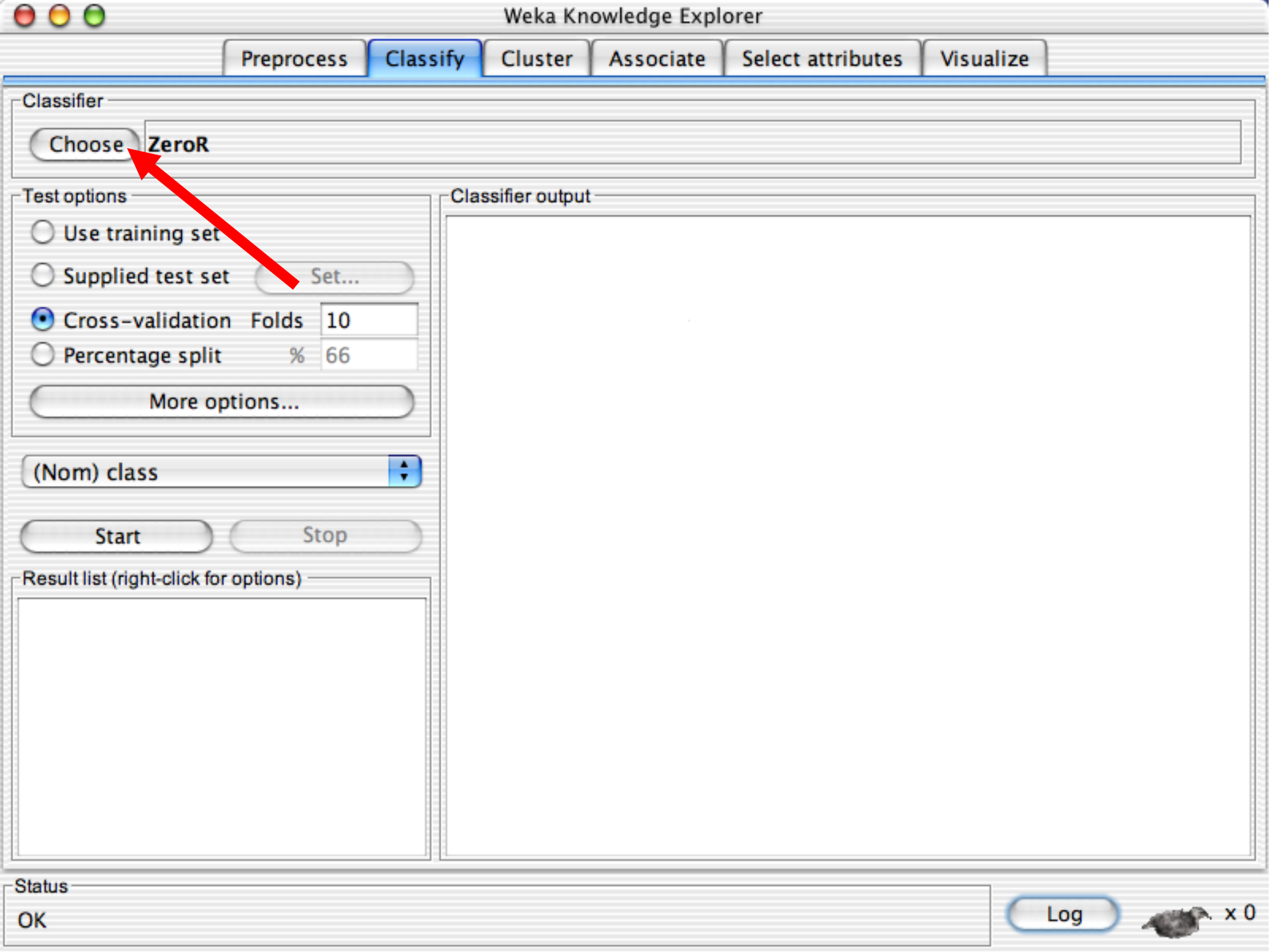
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

ZeroR

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)

Classifier output

Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

- weka
  - classifiers
    - bayes
    - functions
    - lazy
    - meta
    - misc
    - trees
      - adtree
        - DecisionStump
        - Id3
      - j48
        - J48
      - lmt
      - m5
        - RandomForest
        - RandomTree
        - REPTree
        - UserClassifier
    - rules

Classifier output

Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

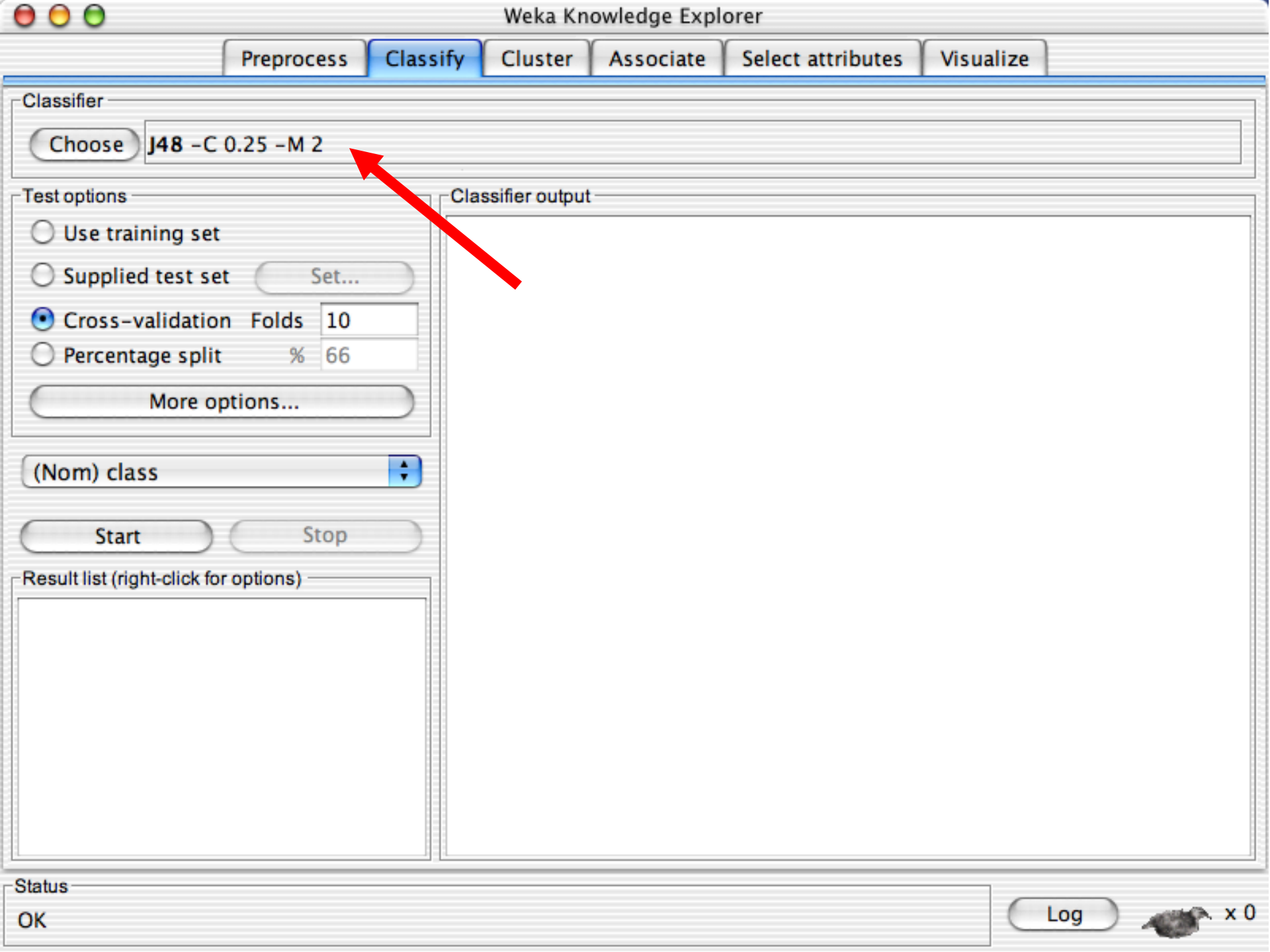
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

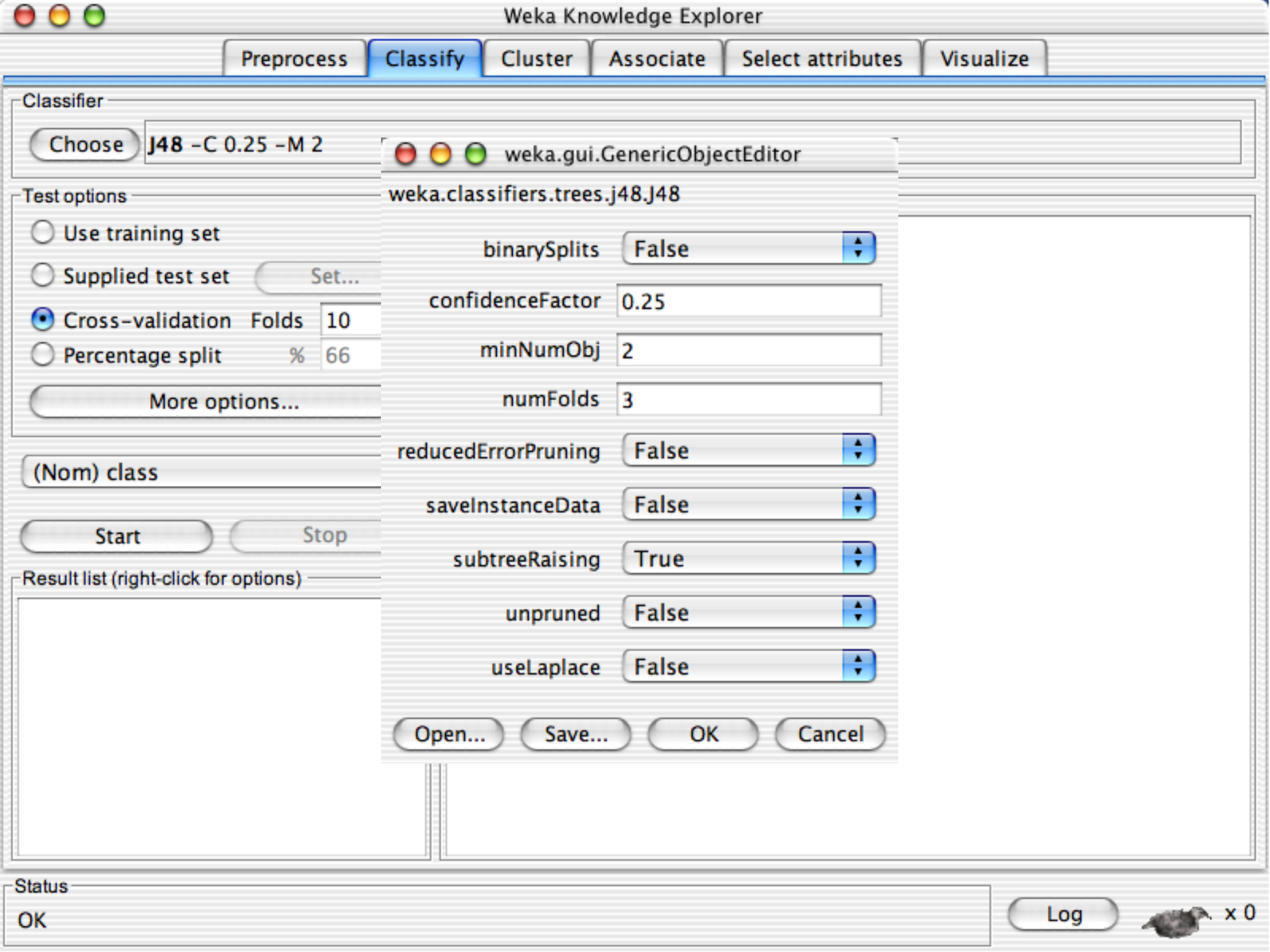
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

weka.gui.GenericObjectEditor

weka.classifiers.trees.j48.J48

binarySplits False

confidenceFactor 0.25

minNumObj 2

numFolds 3

reducedErrorPruning False

saveInstanceData False

subtreeRaising True

unpruned False

useLaplace False

Open...

Save...

OK

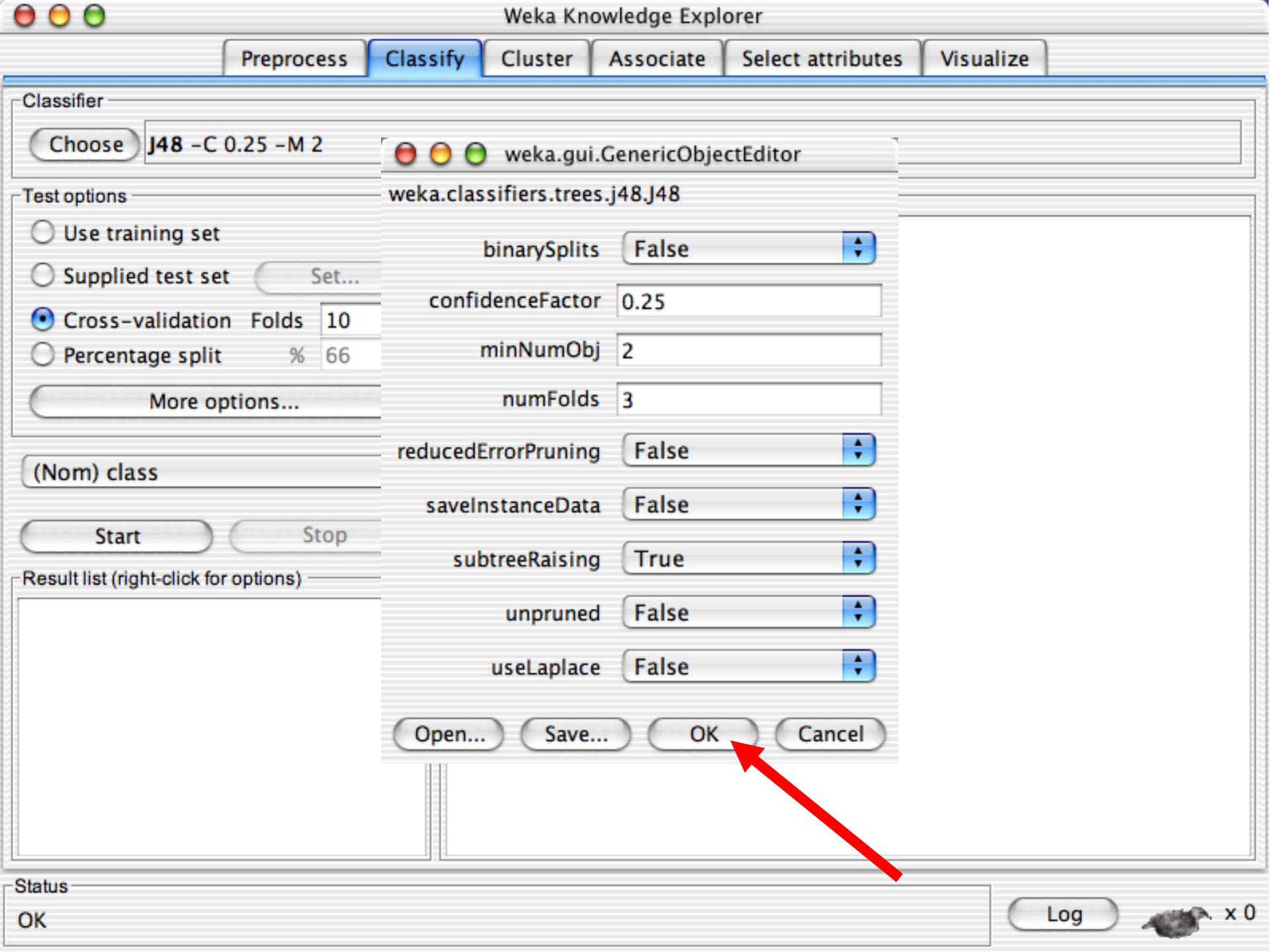
Cancel

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

weka.gui.GenericObjectEditor

weka.classifiers.trees.j48.J48

binarySplits False

confidenceFactor 0.25

minNumObj 2

numFolds 3

reducedErrorPruning False

saveInstanceData False

subtreeRaising True

unpruned False

useLaplace False

Open...

Save...

OK

Cancel

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

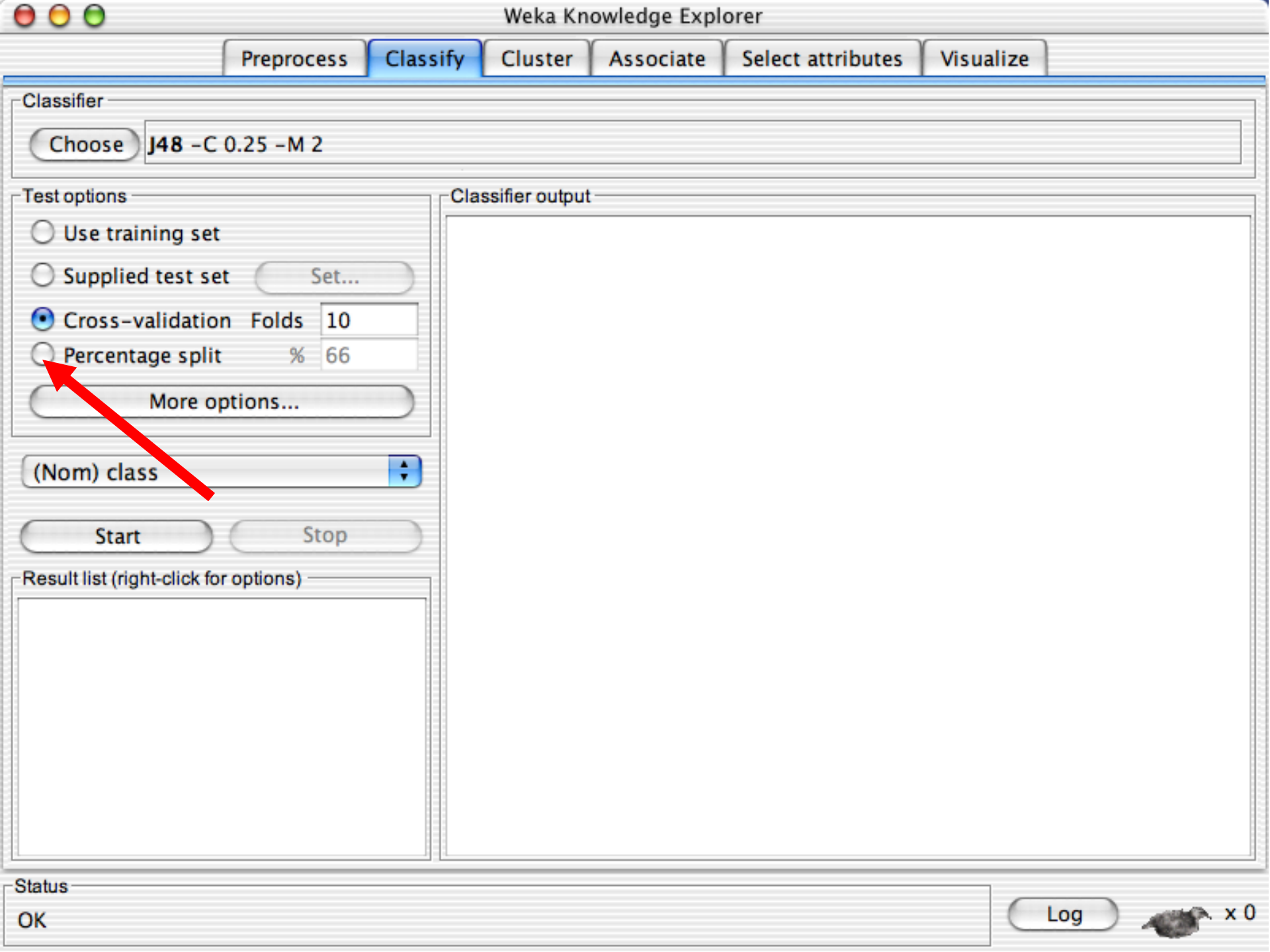
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

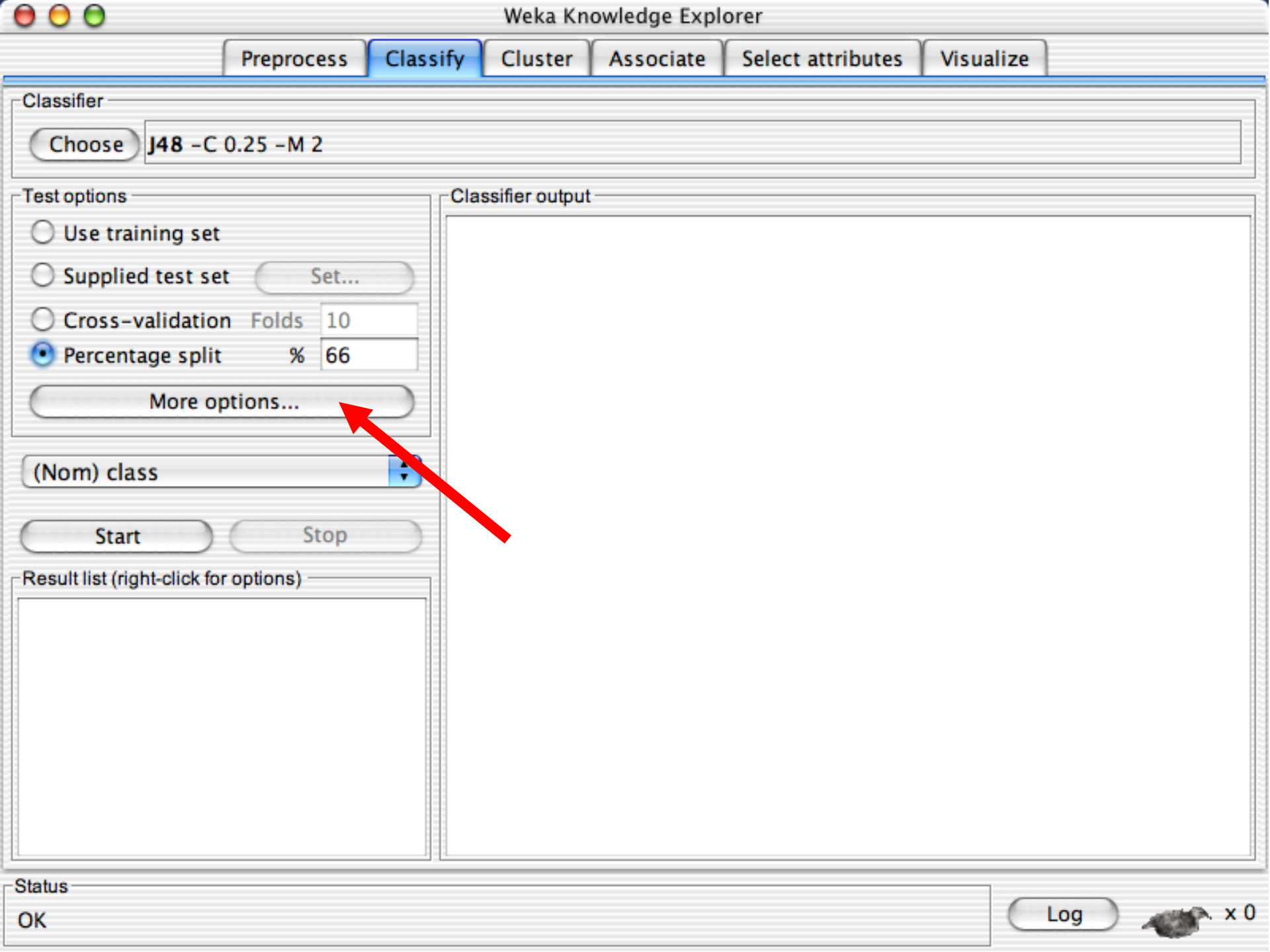
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output



Classifier evaluation opt

☒ Output model☒ Output per-class stats☐ Output entropy evaluation measures☒ Output confusion matrix☒ Store predictions for visualization☐ Output text predictions on test set☐ Cost-sensitive evaluation

Set...

Random seed for XVal / % Split

1

OK

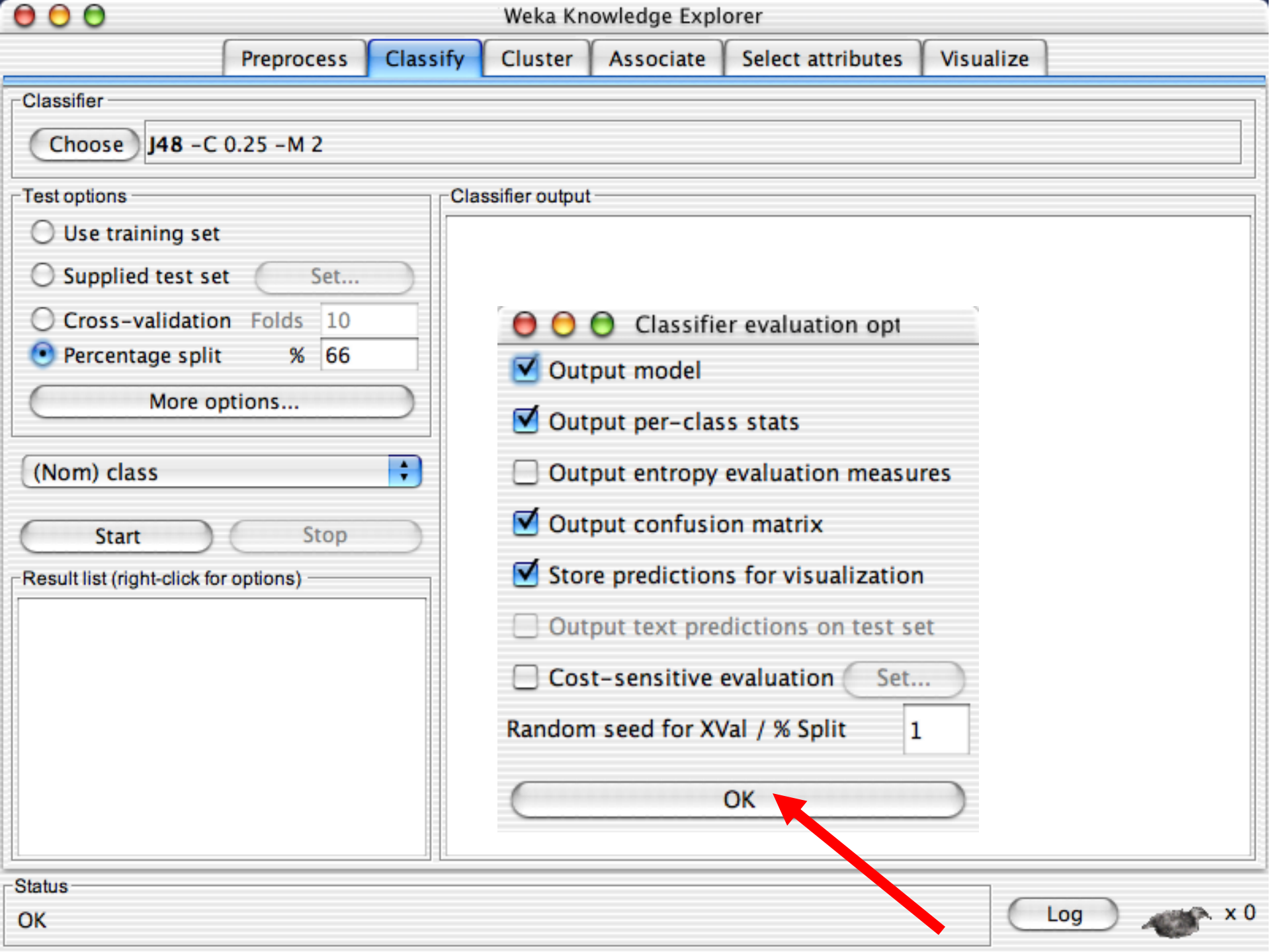
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

Classifier evaluation opt

☒ Output model☒ Output per-class stats☐ Output entropy evaluation measures☒ Output confusion matrix☒ Store predictions for visualization☐ Output text predictions on test set☐ Cost-sensitive evaluation

Set...

Random seed for XVal / % Split

1

OK

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

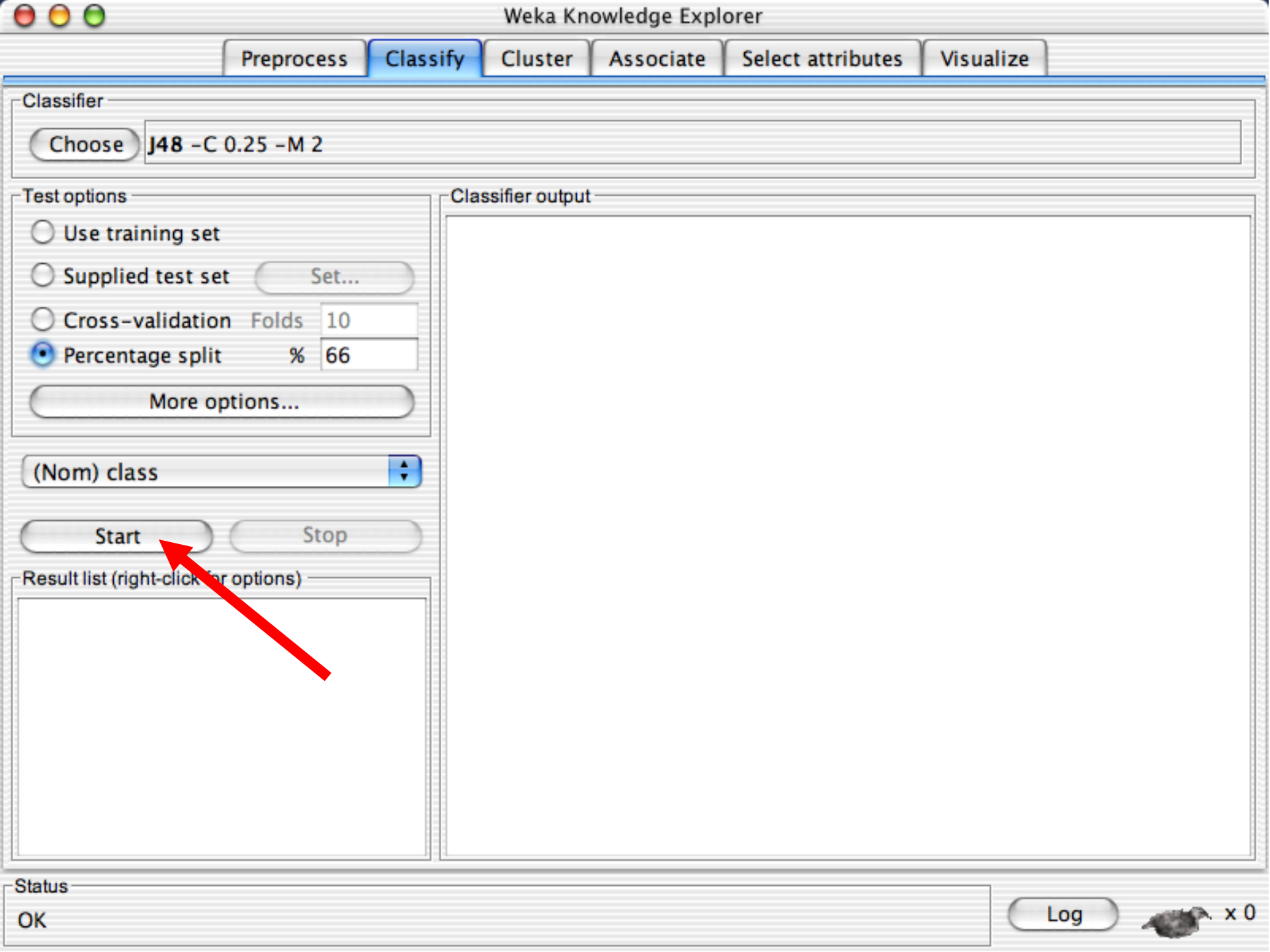
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

Classifier output

Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

Classifier output

=== Run information ===

Scheme: weka.classifiers.trees.j48.J48 -C 0.25 -M 2  
Relation: iris  
Instances: 150  
Attributes: 5

sepalength  
sepalwidth  
petallength  
petalwidth  
class

Test mode: split 66% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

```
-----  
petalwidth <= 0.6: Iris-setosa (50.0)  
petalwidth > 0.6  
|   petalwidth <= 1.7  
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)  
|   |   petallength > 4.9  
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)  
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)  
|   petalwidth > 1.7: Iris-virginica (46.0/1.0)
```

Number of Leaves : 5

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

Classifier output

=== Run information ===

Scheme: weka.classifiers.trees.j48.J48 -C 0.25 -M 2  
Relation: iris  
Instances: 150  
Attributes: 5

sepalength  
sepalwidth  
petallength  
petalwidth  
class

Test mode: split 66% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

```
-----  
petalwidth <= 0.6: Iris-setosa (50.0)  
petalwidth > 0.6  
| petalwidth <= 1.7  
| | petallength <= 4.9: Iris-versicolor (48.0/1.0)  
| | petallength > 4.9  
| | | petalwidth <= 1.5: Iris-virginica (3.0)  
| | | petalwidth > 1.5: Iris-versicolor (3.0/1.0)  
| petalwidth > 1.7: Iris-virginica (46.0/1.0)
```

Number of Leaves : 5



Status

OK

Log

 x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

View in main window

View in separate window

Save result buffer

Load model

Save model

Re-evaluate model on current test set

Visualize classifier errors

Visualize tree

Visualize margin curve

Visualize threshold curve

Visualize cost curve

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| Recall | F-Measure | Class           |
|--------|-----------|-----------------|
| 1      | 1         | Iris-setosa     |
| 1      | 0.95      | Iris-versicolor |
| 0.882  | 0.938     | Iris-virginica  |

Status

OK

Log

x 0

Weka Knowledge Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 - C 0.25 - M 2

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation

☒ Percentage split

More options

(Nom) class

Start

Result list (right-click for details)

11:49:05 - trees.j48.J48

Weka Classifier Tree Visualizer: 11:49:05 - trees.j48.J48 (iris)

Tree View

petalwidth

<= 0.6

> 0.6

Iris-setosa (50.0)

petalwidth

<= 1.7

> 1.7

petallength

<= 4.9

> 4.9

Iris-versicolor (48.0/1.0)

petalwidth

<= 1.5

> 1.5

Iris-virginica (3.0)

Iris-versicolor (3.0/1.0)

Iris-virginica (46.0/1.0)

96.0784 %

3.9216 %

ass

is-setosa

is-versicolor

is-virginica

15 0 0 | a = Iris-setosa

0 19 0 | b = Iris-versicolor

0 2 15 | c = Iris-virginica

Status

OK

Log

 x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

View in main window

View in separate window

Save result buffer

Load model

Save model

Re-evaluate model on current test set

Visualize classifier errors

Visualize tree

Visualize margin curve

Visualize threshold curve

Visualize cost curve

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| Recall | F-Measure | Class           |
|--------|-----------|-----------------|
| 1      | 1         | Iris-setosa     |
| 1      | 0.95      | Iris-versicolor |
| 0.882  | 0.938     | Iris-virginica  |

lor  
ca

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☐ Cross-validation
- ☒ Percentage split

X: petallength (Num)

Y: petalwidth (Num)

Colour: class (Nom)

Select Instance

Reset

Clear

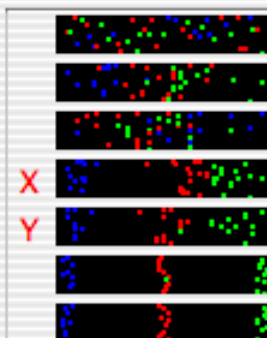
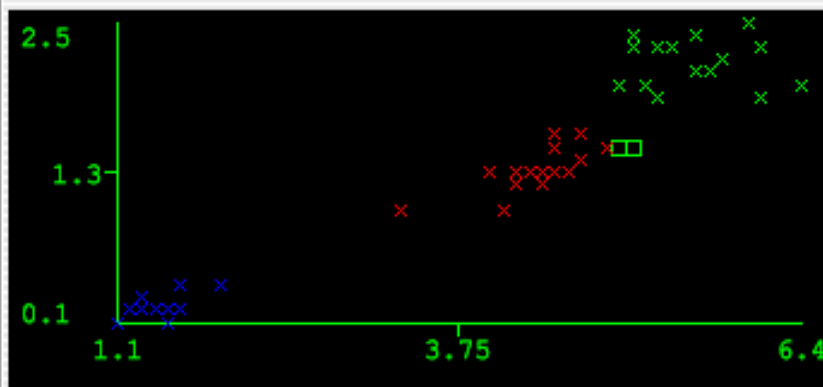
Save

Jitter

96.0784 %

3.9216 %

More options Plot: iris\_predicted



Class colour

Iris-setosa Iris-versicolor Iris-virginica

```
0 15 0 | D = Iris-versicolor
0 2 15 | c = Iris-virginica
```

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

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)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Classifier

- weka
  - classifiers
    - bayes
    - functions
      - LeastMedSq
      - LinearRegression
      - Logistic
      - neural
        - NeuralNetwork
      - pace
      - supportVector
        - SimpleLinearRegression
        - SimpleLogistic
        - VotedPerceptron
        - Winnow
    - lazy
    - meta
    - misc
    - trees
    - rules

## output

Time taken to build model: 0.24 seconds

Evaluation on test split ===

Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Absolute error                   | 0.0396    |           |
| Average squared error            | 0.1579    |           |
| Weighted average absolute error  | 8.8979 %  |           |
| Relative squared error           | 33.4091 % |           |
| Number of Instances              | 51        |           |

Failed Accuracy By Class ===

|  | FP Rate | Precision | Recall | F-Measure | Class           |
|--|---------|-----------|--------|-----------|-----------------|
|  | 0       | 1         | 1      | 1         | Iris-setosa     |
|  | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
|  | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

Confusion Matrix ===

```
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
```

Status

OK

Log



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

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)

11:49:05 - trees.j48.J48

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

```

a  b  c  <-- classified as
15  0  0 | a = Iris-setosa
 0 19  0 | b = Iris-versicolor
 0  2 15 | c = Iris-virginica

```

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

**Choose** NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

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)

11:49:05 - trees.j48.J48

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 49        | 96.0784 % |
| Incorrectly Classified Instances | 2         | 3.9216 %  |
| Kappa statistic                  | 0.9408    |           |
| Mean absolute error              | 0.0396    |           |
| Root mean squared error          | 0.1579    |           |
| Relative absolute error          | 8.8979 %  |           |
| Root relative squared error      | 33.4091 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

|    |    |    |                     |
|----|----|----|---------------------|
| a  | b  | c  | <-- classified as   |
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0

Preprocess

Classify

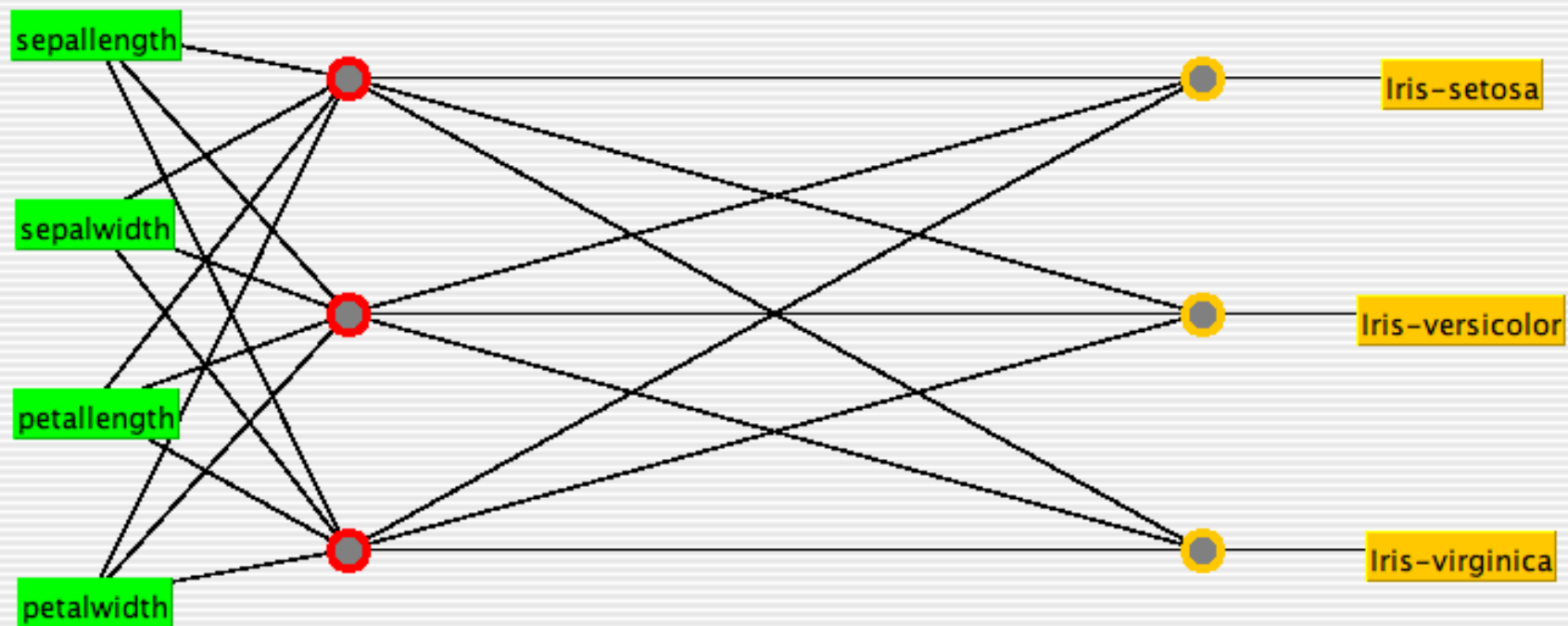
Cluster

Associate

Select attributes

Visualize

Neural Network



Controls

Start

Epoch 0

Num Of Epochs 500

Accept

Error per Epoch = 0

Learning Rate = 0.3

Momentum = 0.2

building model on training data...

Preprocess

Classify

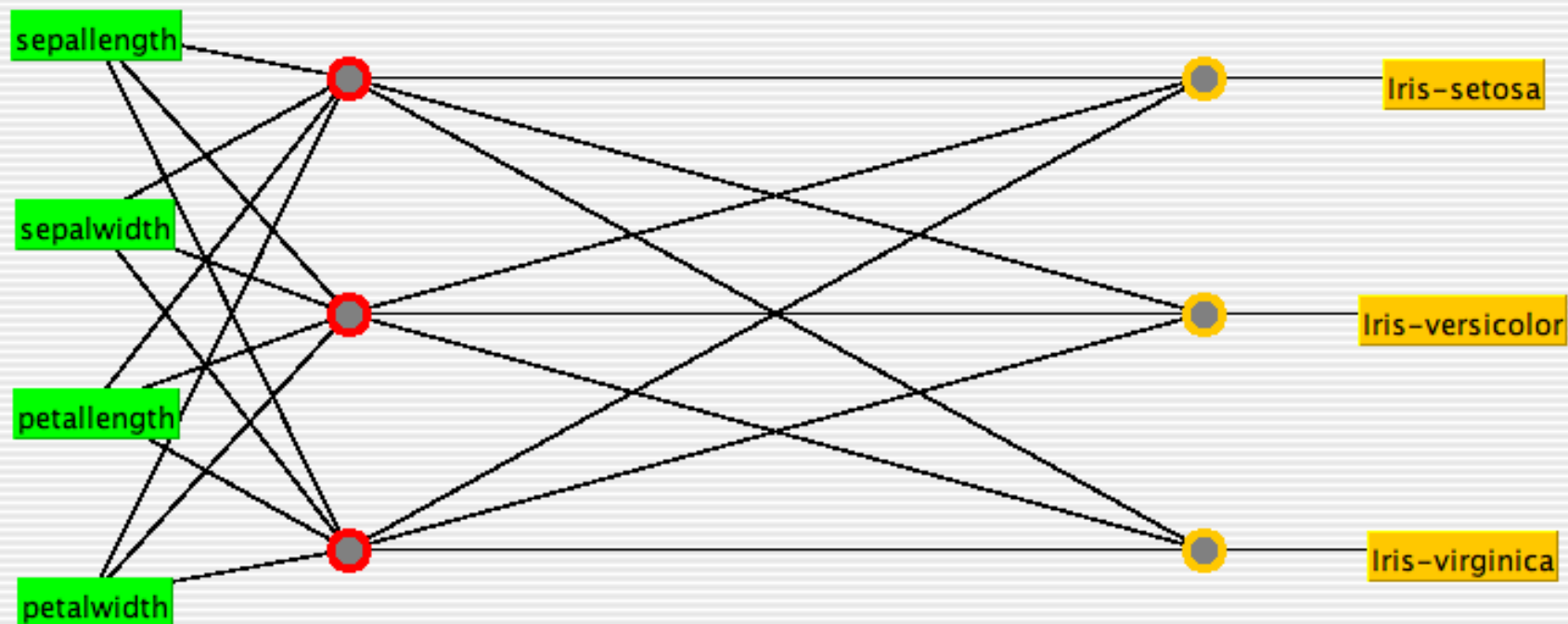
Cluster

Associate

Select attributes

Visualize

Neural Network



Controls

Start

Epoch 0

Num Of Epochs 500

Accept

Error per Epoch = 0

Learning Rate = 0.3

Momentum = 0.2

building model on training data...

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 50        | 98.0392 % |
| Incorrectly Classified Instances | 1         | 1.9608 %  |
| Kappa statistic                  | 0.9704    |           |
| Mean absolute error              | 0.0239    |           |
| Root mean squared error          | 0.1101    |           |
| Relative absolute error          | 5.3594 %  |           |
| Root relative squared error      | 23.2952 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.031   | 0.95      | 1      | 0.974     | Iris-versicolor |
| 0.941   | 0       | 1         | 0.941  | 0.97      | Iris-virginica  |

=== Confusion Matrix ===

```

a  b  c  <-- classified as
15  0  0 | a = Iris-setosa
 0 19  0 | b = Iris-versicolor
 0  1 16 | c = Iris-virginica

```

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 50        | 98.0392 % |
| Incorrectly Classified Instances | 1         | 1.9608 %  |
| Kappa statistic                  | 0.9704    |           |
| Mean absolute error              | 0.0239    |           |
| Root mean squared error          | 0.1101    |           |
| Relative absolute error          | 5.3594 %  |           |
| Root relative squared error      | 23.2952 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.031   | 0.95      | 1      | 0.974     | Iris-versicolor |
| 0.941   | 0       | 1         | 0.941  | 0.97      | Iris-virginica  |

=== Confusion Matrix ===

```

a b c  <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 1 16 | c = Iris-virginica

```

Status

OK

Log

x 0

Weka Knowledge Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

weka

classifiers

bayes

AODE

BayesNetK2

BayesNetB

NaiveBayes

NaiveBayesMultinomial

NaiveBayesSimple

NaiveBayesUpdateable

functions

lazy

meta

misc

trees

rules

Classifier output

== Evaluation on test split ==

== Summary ==

Correctly Classified Instances

50

98.0392 %

Incorrectly Classified Instances

1

1.9608 %

Nappa statistic

0.9704

Mean absolute error

0.0239

Root mean squared error

0.1101

Relative absolute error

5.3594 %

Root relative squared error

23.2952 %

Total Number of Instances

51

== Detailed Accuracy By Class ==

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.031   | 0.95      | 1      | 0.974     | Iris-versicolor |
| 0.941   | 0       | 1         | 0.941  | 0.97      | Iris-virginica  |

== Confusion Matrix ==

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 1  | 16 | c = Iris-virginica  |

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Log

 x 0

Weka Knowledge Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

NaiveBayes

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

Folds

10

%

66

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 50        | 98.0392 % |
| Incorrectly Classified Instances | 1         | 1.9608 %  |
| Kappa statistic                  | 0.9704    |           |
| Mean absolute error              | 0.0239    |           |
| Root mean squared error          | 0.1101    |           |
| Relative absolute error          | 5.3594 %  |           |
| Root relative squared error      | 23.2952 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.031   | 0.95      | 1      | 0.974     | Iris-versicolor |
| 0.941   | 0       | 1         | 0.941  | 0.97      | Iris-virginica  |

=== Confusion Matrix ===

a

b

c

<-- classified as

15

0

0

| a = Iris-setosa

0

19

0

| b = Iris-versicolor

0

1

16

| c = Iris-virginica

Log

 x 0

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Weka Knowledge Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

NaiveBayes

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

Folds

10

%

66

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

Classifier output

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances

50

98.0392 %

Incorrectly Classified Instances

1

1.9608 %

Kappa statistic

0.9704

Mean absolute error

0.0239

Root mean squared error

0.1101

Relative absolute error

5.3594 %

Root relative squared error

23.2952 %

Total Number of Instances

51

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.031   | 0.95      | 1      | 0.974     | Iris-versicolor |
| 0.941   | 0       | 1         | 0.941  | 0.97      | Iris-virginica  |

=== Confusion Matrix ===

a

b

c

<-- classified as

15

0

0

| a = Iris-setosa

0

19

0

| b = Iris-versicolor

0

1

16

| c = Iris-virginica

Log

 x 0

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NaiveBayes

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Mean absolute error              | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947   | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882   | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

=== Confusion Matrix ===

```

a b c  <-- classified as
15 0 0 | a = Iris-setosa
0 18 1 | b = Iris-versicolor
0 2 15 | c = Iris-virginica

```

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NaiveBayes

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Mean absolute error              | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947   | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882   | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

=== Confusion Matrix ===

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 18 | 1  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NaiveBayes

Test options

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

More options...

(Nom) class

Start

Result list (right-click for)

11:49:05 - trees.j48.J

14:34:28 - functions.

14:48:05 - bayes.Nai

View in main window

View in separate window

Save result buffer

Load model

Save model

Re-evaluate model on current test set

Visualize classifier errors

Visualize tree

Visualize margin curve

Visualize threshold curve

Visualize cost curve

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Mean absolute error              | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

|  | Precision | Recall | F-Measure | Class           |
|--|-----------|--------|-----------|-----------------|
|  | 1         | 1      | 1         | Iris-setosa     |
|  | 0.9       | 0.947  | 0.923     | Iris-versicolor |
|  | 0.938     | 0.882  | 0.909     | Iris-virginica  |

x ===

classified as  
Iris-setosa  
Iris-versicolor  
Iris-virginica

Iris-setosa

Iris-versicolor

Iris-virginica

Status

OK

Log



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set☐ Supplied test set☐ Cross-validation For☒ Percentage split

More options

(Nom) class

Start

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neu

14:48:05 - bayes.NaiveBa

Weka Classifier Visualize: ThresholdCurve. Class value Iris-versicolor)

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

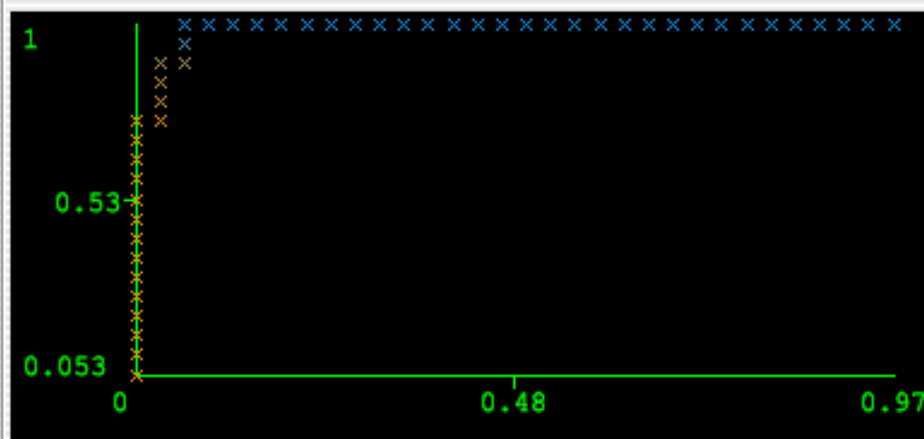
Reset

Clear

Save

Jitter

Plot: ThresholdCurve



Class colour

0

0.5

1

X

Y

osa  
sicolor  
ynica

Log

x 0

Preprocess

**Classify**

Cluster

Associate

Select attributes

Visualize

Classifier

Choose **NaiveBayes**

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Mean absolute error              | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947   | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882   | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

=== Confusion Matrix ===

```

a b c  <-- classified as
15 0 0 | a = Iris-setosa
0 18 1 | b = Iris-versicolor
0 2 15 | c = Iris-virginica

```

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose NaiveBayes

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Mean absolute error              | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947   | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882   | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

=== Confusion Matrix ===

```
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 18 1 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
```

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

weka

- classifiers
  - bayes
  - functions
  - lazy
  - meta
  - misc
  - trees
    - adtree
      - DecisionStump
      - Id3
    - j48
    - lmt
    - m5
      - RandomForest
      - RandomTree
      - REPTree
      - UserClassifier
  - rules

Classifier output

== Evaluation on test split ==

== Summary ==

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Average absolute error           | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

== Detailed Accuracy By Class ==

| P Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|--------|---------|-----------|--------|-----------|-----------------|
| 1      | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947  | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882  | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

== Confusion Matrix ==

| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 18 | 1  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

QuickTime™ and a TIFF (LZW) decompressor are needed to s

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)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 48        | 94.1176 % |
| Incorrectly Classified Instances | 3         | 5.8824 %  |
| Kappa statistic                  | 0.9113    |           |
| Mean absolute error              | 0.0447    |           |
| Root mean squared error          | 0.1722    |           |
| Relative absolute error          | 10.0365 % |           |
| Root relative squared error      | 36.4196 % |           |
| Total Number of Instances        | 51        |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947   | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882   | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

=== Confusion Matrix ===

|    |    |    |                     |
|----|----|----|---------------------|
| a  | b  | c  | <-- classified as   |
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 18 | 1  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set☐ Supplied test set☐ Cross-validation For☒ Percentage split

More options

(Nom) class

Start

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neu

14:48:05 - bayes.NaiveBa

15:26:57 - trees.UserClas

Tree Visualizer

Data Visualizer

Tree View

[Iris-setosa, 50.0]  
[Iris-versicolor, 50.0]  
[Iris-virginica, 50.0]

Status

Building model on training data...

Log



x 1

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set☐ Supplied test set☐ Cross-validation☒ Percentage split

More...

(Nom) class

Start

Result list (right-click for)

11:49:05 - trees.j48.

14:34:28 - functions

14:48:05 - bayes.Naï

15:26:57 - trees.Use



Tree Visualizer

Data Visualizer

X: petallength (Num)

Y: petalwidth (Num)

Colour: class (Nom)

Polyline

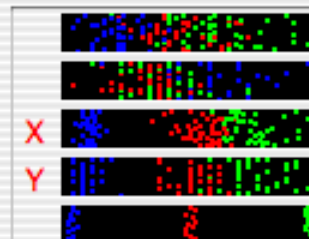
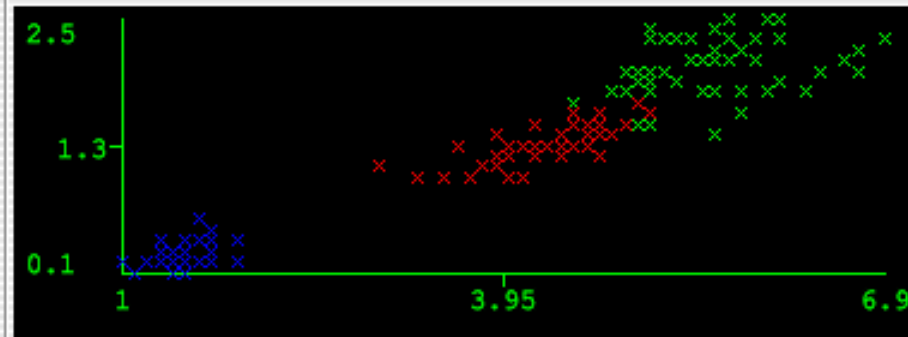
Submit

Clear

Save

Jitter

Plot: iris



Class colour

Iris-setosa Iris-versicolor Iris-virginica

Status

Building model on training data...

Log



x 1

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set☐ Supplied test set☐ Cross-validation☒ Percentage split

More...

(Nom) class

Start

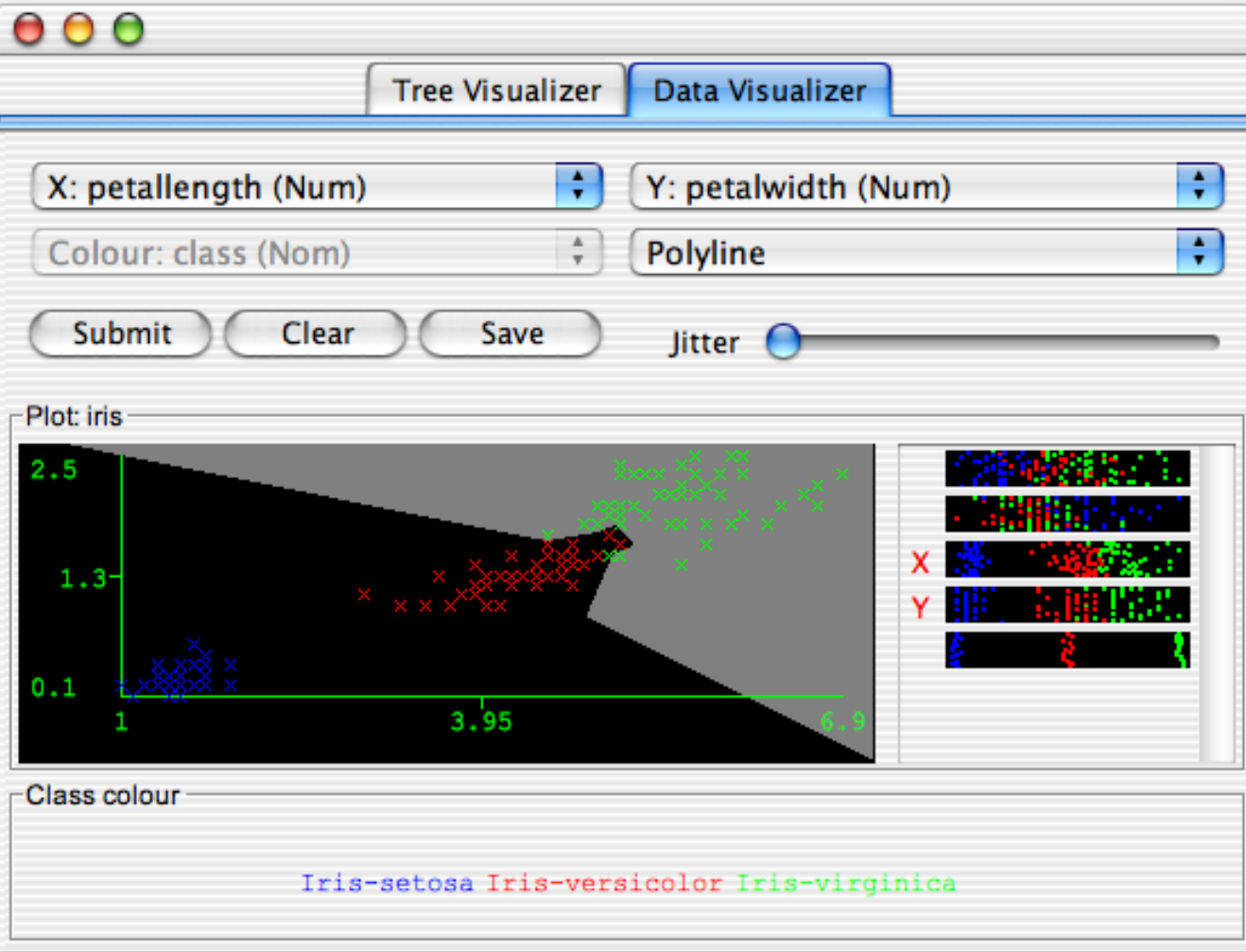
Result list (right-click for)

11:49:05 - trees.j48.

14:34:28 - functions

14:48:05 - bayes.Naï

15:26:57 - trees.Use



Status

Building model on training data...

Log



x 1

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set☐ Supplied test set☐ Cross-validation☒ Percentage split

More options

(Nom) class

Start

Result list (right-click for context menu)

11:49:05 - trees.j48.J

14:34:28 - functions.

14:48:05 - bayes.Naïve

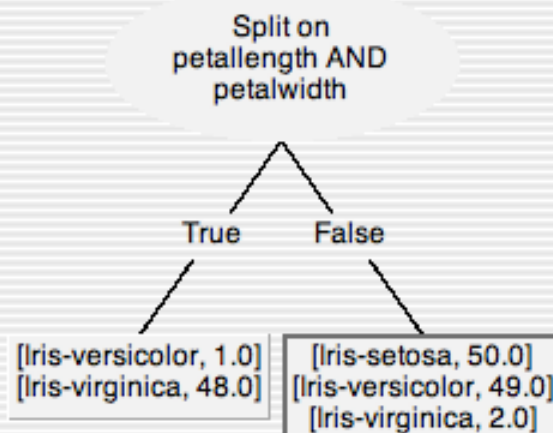
15:26:57 - trees.UserClassifier



Tree Visualizer

Data Visualizer

Tree View



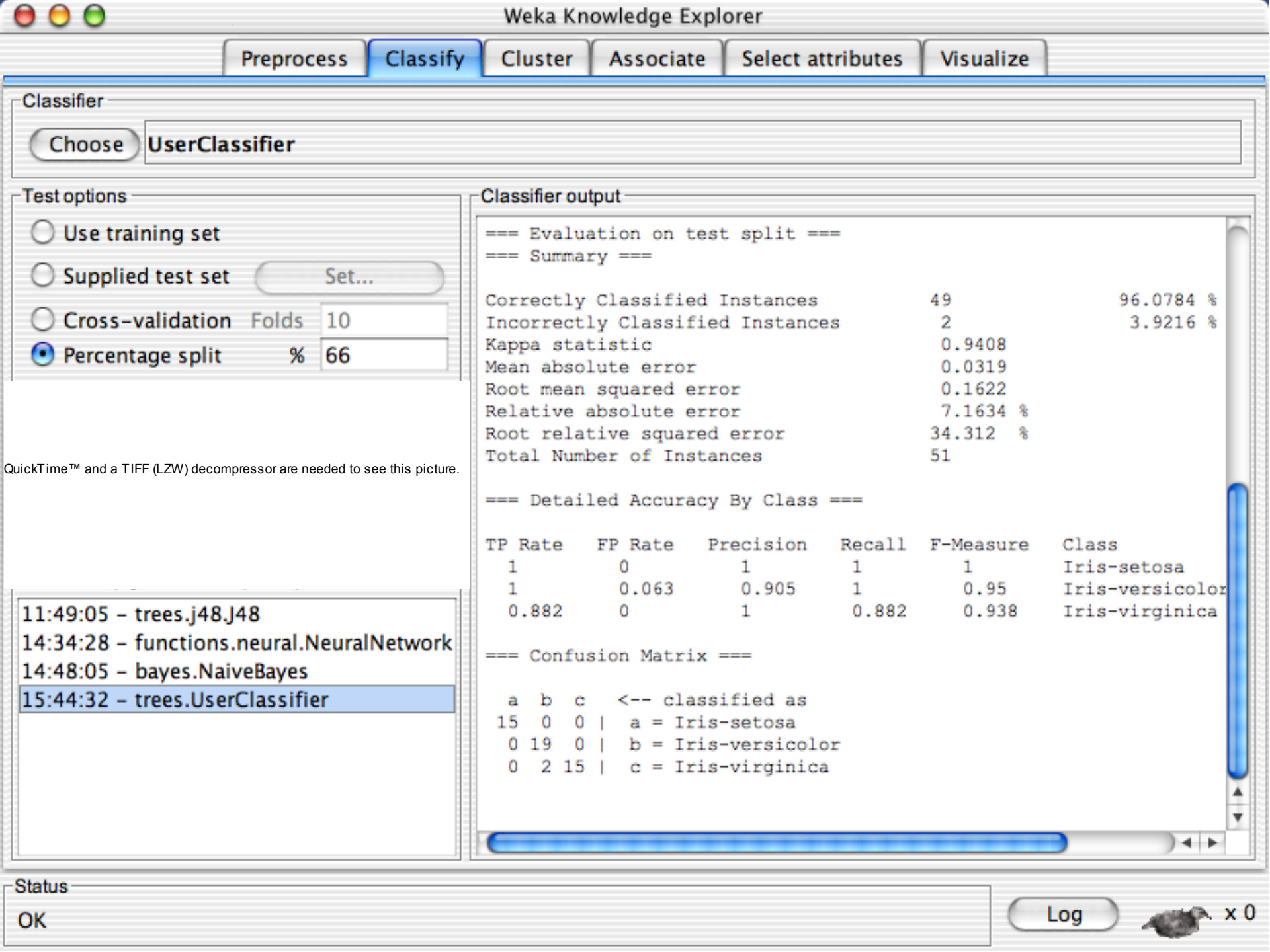
Status

Building model on training data...

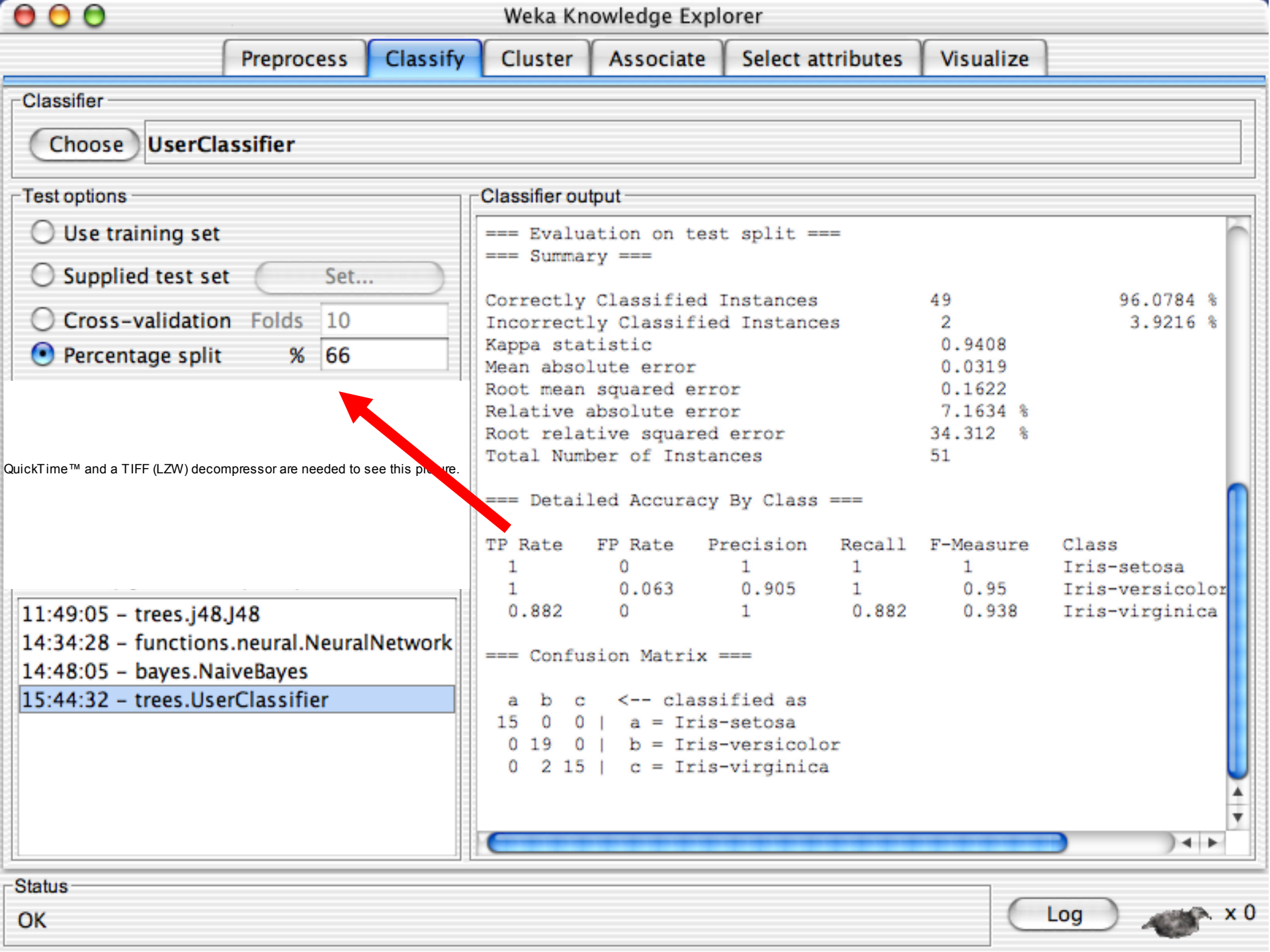
Log



x 1



QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set☐ Supplied test set

Set...

☐ Cross-validation

Folds

10

☒ Percentage split

%

66

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |          |           |
|----------------------------------|----------|-----------|
| Correctly Classified Instances   | 49       | 96.0784 % |
| Incorrectly Classified Instances | 2        | 3.9216 %  |
| Kappa statistic                  | 0.9408   |           |
| Mean absolute error              | 0.0319   |           |
| Root mean squared error          | 0.1622   |           |
| Relative absolute error          | 7.1634 % |           |
| Root relative squared error      | 34.312 % |           |
| Total Number of Instances        | 51       |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

|    |    |    |                     |
|----|----|----|---------------------|
| a  | b  | c  | <-- classified as   |
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 19 | 0  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

15:44:32 - trees.UserClassifier

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set☐ Supplied test set

Set...

☐ Cross-validation Folds 10☒ Percentage split % 66

(Num) sepallength

(Num) sepalwidth

✓ (Num) petallength

(Num) petalwidth

(Nom) class

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

15:44:32 - trees.UserClassifier

Classifier output

=== Evaluation on test split ===

=== Summary ===

|                                  |          |           |
|----------------------------------|----------|-----------|
| Correctly Classified Instances   | 49       | 96.0784 % |
| Incorrectly Classified Instances | 2        | 3.9216 %  |
| Kappa statistic                  | 0.9408   |           |
| Mean absolute error              | 0.0319   |           |
| Root mean squared error          | 0.1622   |           |
| Relative absolute error          | 7.1634 % |           |
| Root relative squared error      | 34.312 % |           |
| Total Number of Instances        | 51       |           |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

```

a  b  c  <-- classified as
15  0  0 | a = Iris-setosa
 0 19  0 | b = Iris-versicolor
 0  2 15 | c = Iris-virginica

```

Status

OK

Log

x 0

Preprocess

**Classify**

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

UserClassifier

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation

☒ Percentage split

Set...

Folds

%

10

66

Classifier output

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances

Incorrectly Classified Instances

Kappa statistic

Mean absolute error

Root mean squared error

Relative absolute error

Root relative squared error

Total Number of Instances

49

2

0.9408

0.0319

0.1622

7.1634 %

34.312 %

51

96.0784 %

3.9216 %

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

=== Confusion Matrix ===

a

b

c

<-- classified as

15

0

0

| a = Iris-setosa

0

19

0

| b = Iris-versicolor

0

2

15

| c = Iris-virginica

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

15:44:32 - trees.UserClassifier

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Classifier

- weka
  - classifiers
    - bayes
    - functions
    - lazy
    - meta
    - misc
    - trees
      - adtree
      - DecisionStump
      - Id3
      - j48
      - lmt
      - m5
        - M5P**
        - RandomForest
        - RandomTree
        - REPTree
        - UserClassifier
  - rules

## Classifier output

== Evaluation on test split ==

== Summary ==

```

Correctly Classified Instances      49           96.0784 %
Incorrectly Classified Instances     2           3.9216 %
Kappa statistic                     0.9408
Mean absolute error                  0.0319
Root mean squared error              0.1622
Relative absolute error              7.1634 %
Root relative squared error         34.312 %
Total Number of Instances           51

```

== Detailed Accuracy By Class ==

| TP Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|---------|---------|-----------|--------|-----------|-----------------|
| 1       | 0       | 1         | 1      | 1         | Iris-setosa     |
| 1       | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| 0.882   | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

== Confusion Matrix ==

```

a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica

```

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose M5P -M 4.0

Test options

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

More options...

(Num) petallength

Start

Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

15:44:32 - trees.UserClassifier

15:49:03 - trees.m5.M5P

Classifier output

=== Run information ===

Scheme: weka.classifiers.trees.m5.M5P -M 4.0

Relation: iris

Instances: 150

Attributes: 5

sepalength

sepalwidth

petallength

petalwidth

class

Test mode: split 66% train, remainder test

=== Classifier model (full training set) ===

M5 pruned model tree:

(using smoothed predictions)

petalwidth &lt;= 0.8 : LM1 (50/10.469%)

petalwidth &gt; 0.8 :

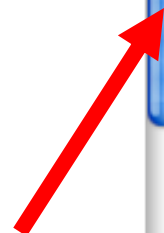
| class=Iris-virginica &lt;= 0.5 : LM2 (50/14.325%)

| class=Iris-virginica &gt; 0.5 : LM3 (50/17.598%)

LM num: 1

Linear Regression Model

petallength =

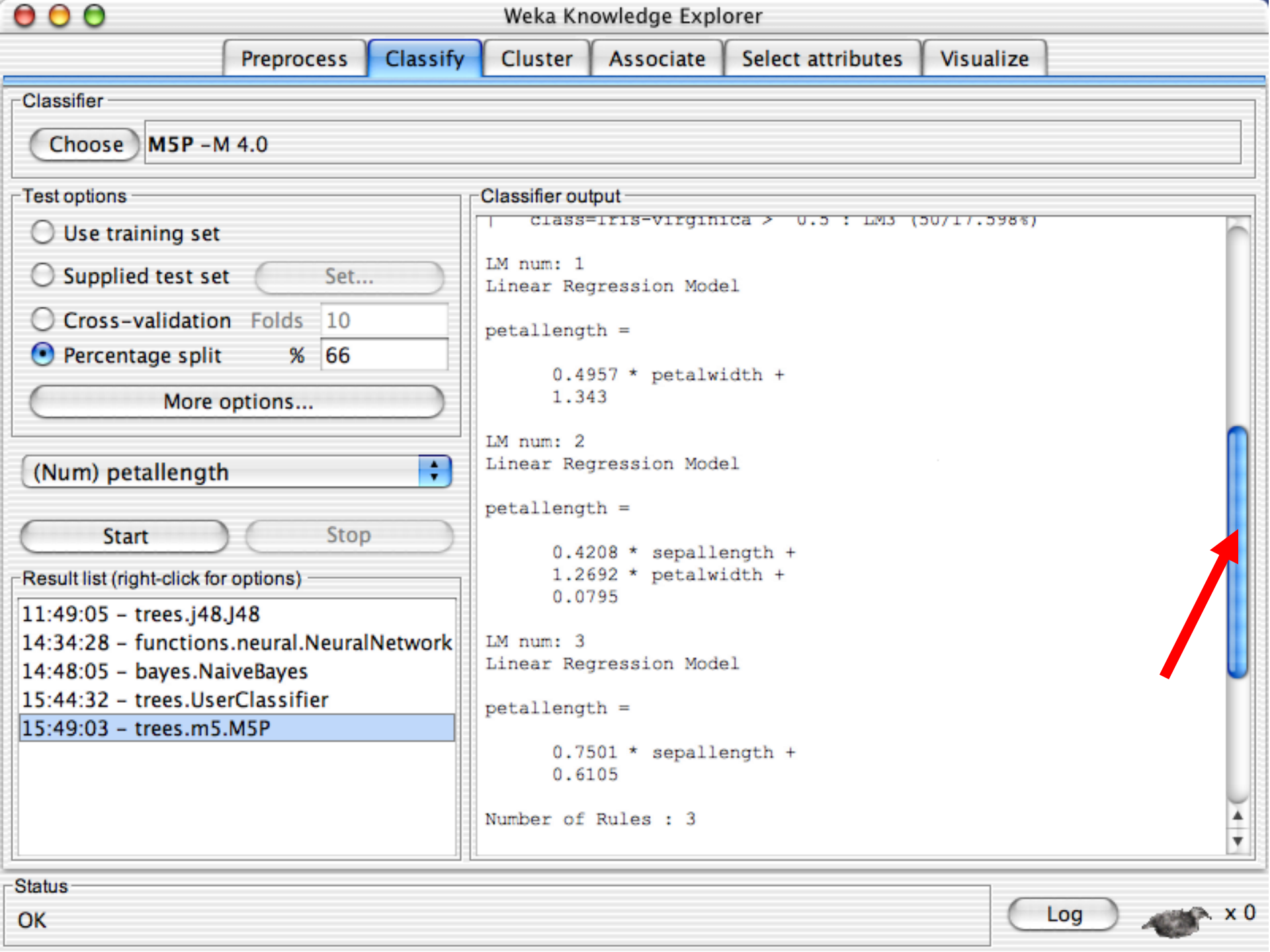


Status

OK

Log





Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose M5P - M 4.0

Test options

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

More options...

(Num) petallength

Start

Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes  
15:44:32 - trees.UserClassifier  
15:49:03 - trees.m5.M5P

Classifier output

class=iris-virginica &gt; 0.5 : LM3 (50/17.598%)

LM num: 1  
Linear Regression Model

petallength =  
  
0.4957 \* petalwidth +  
1.343

LM num: 2  
Linear Regression Model

petallength =  
  
0.4208 \* sepallength +  
1.2692 \* petalwidth +  
0.0795

LM num: 3  
Linear Regression Model

petallength =  
  
0.7501 \* sepallength +  
0.6105

Number of Rules : 3

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

M5P -M 4.0

Test options

☐ Use training set☐ Supplied test set

Set...

☐ Cross-validation Folds 10☒ Percentage split % 66

More options...

(Num) petallength



Start

Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

15:44:32 - trees.UserClassifier

15:49:03 - trees.m5.M5P

Classifier output

```
0.4208 * sepallength +
1.2692 * petalwidth +
0.0795
```

LM num: 3

Linear Regression Model

petallength =

```
0.7501 * sepallength +
0.6105
```

Number of Rules : 3

Time taken to build model: 1.31 seconds

=== Evaluation on test split ===

=== Summary ===

|                             |           |
|-----------------------------|-----------|
| Correlation coefficient     | 0.9889    |
| Mean absolute error         | 0.1861    |
| Root mean squared error     | 0.255     |
| Relative absolute error     | 11.9578 % |
| Root relative squared error | 14.9153 % |
| Total Number of Instances   | 51        |

Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

M5P -M 4.0

Test options

☐ Use training set☐ Supplied test set

Set...

☐ Cross-validation Folds 10☒ Percentage split % 66

More options...

(Num) petallength

Start

Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

15:44:32 - trees.UserClassifier

15:49:03 - trees.m5.M5P

Classifier output

```
0.4208 * sepallength +  
1.2692 * petalwidth +  
0.0795
```

LM num: 3

Linear Regression Model

petallength =

```
0.7501 * sepallength +  
0.6105
```

Number of Rules : 3

Time taken to build model: 1.31 seconds

=== Evaluation on test split ===

=== Summary ===

|                             |           |
|-----------------------------|-----------|
| Correlation coefficient     | 0.9889    |
| Mean absolute error         | 0.1861    |
| Root mean squared error     | 0.255     |
| Relative absolute error     | 11.9578 % |
| Root relative squared error | 14.9153 % |
| Total Number of Instances   | 51        |

Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

M5P

Weka Classifier Visualize: 15:49:03 - trees.m5.M5P (iris)

Test options

☐ Use training☐ Supplied test☐ Cross-validation☒ Percentage split

X: sepalength (Num)

Y: petalwidth (Num)

Colour: petallength (Num)

Select Instance

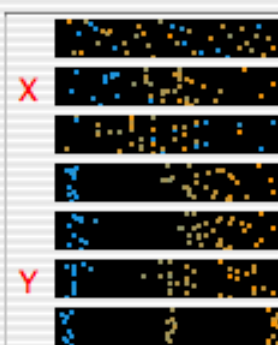
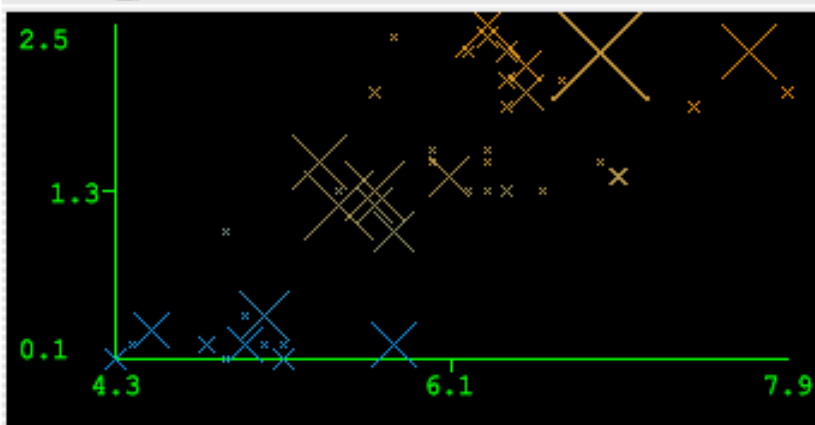
Reset

Clear

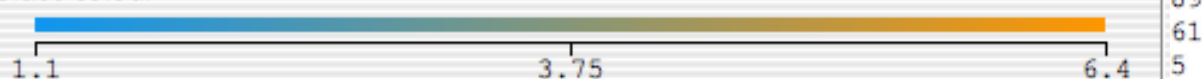
Save

Jitter

Plot: iris\_predicted



Class colour



Root relative squared error

14.9153 %

Total Number of Instances

51

Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

M5P

Weka Classifier Visualize: 15:49:03 - trees.m5.M5P (iris)

Test options

☐ Use training☐ Supplied test☐ Cross-validation☒ Percentage split

X: sepalength (Num)

Y: petalwidth (Num)

Colour: petallength (Num)

Select Instance

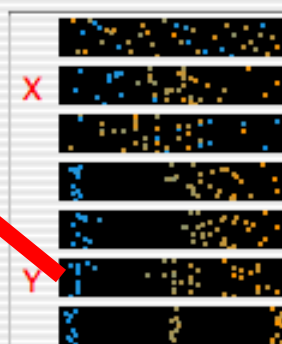
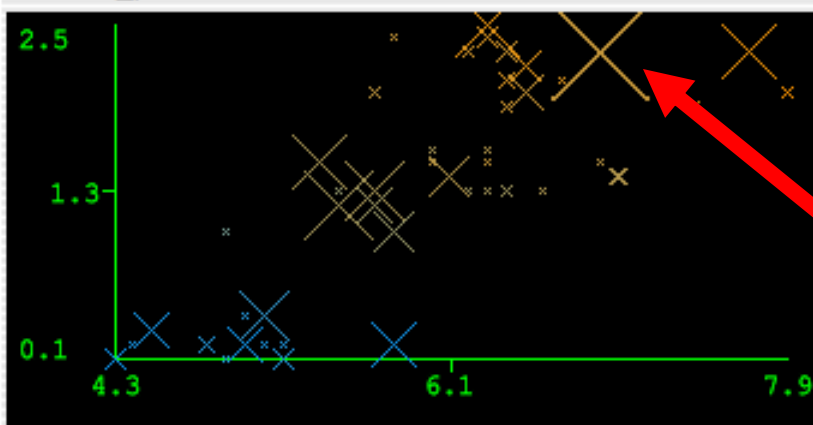
Reset

Clear

Save

Jitter

Plot: iris\_predicted



Class colour

1.1

3.75

6.4

89

61

5

78 %

Root relative squared error

14.9153 %

Total Number of Instances

51

Status

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

M5P

Test options

☐ Use training☐ Supplied test☐ Cross-validation☒ Percentage split

Mo

(Num) petallen

Start

Result list (right-click)

11:49:05 - trees.

14:34:28 - functi

14:48:05 - bayes

15:44:32 - trees.

15:49:03 - trees.

X: sepallength (Num)

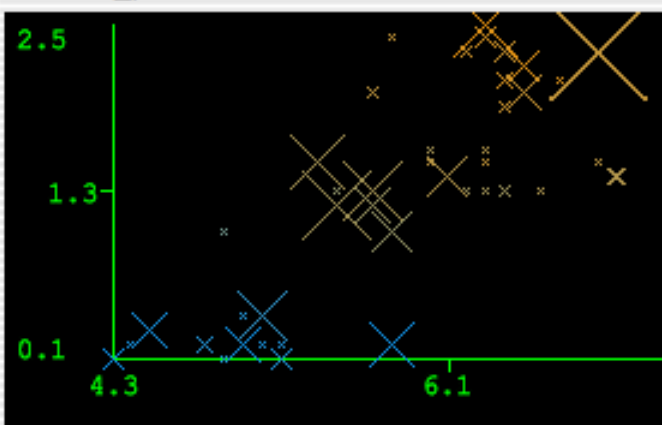
Colour: petallength (Num)

Reset

Clear

Save

Plot: iris\_predicted



Class colour

1.1

3.75

Root relative  
Total Number o

Y:

Sel

Plot : 15:49:03 - trees.m5.M5P (iris)

Instance: 31

Instance\_number : 31.0

sepallength : 6.9

sepalwidth : 3.1

predictedpetallength : 5.892812341943582

petallength : 5.1

petalwidth : 2.3

class : Iris-virginica

Weka : Instance info

Status

OK

Log



x 0

# Explorer: clustering data

- WEKA contains “clusterers” for finding groups of similar instances in a dataset
- Implemented schemes are:
  - ◆ *k*-Means, EM, Cobweb, X-means, FarthestFirst
- Clusters can be visualized and compared to “true” clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a probability distribution



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

EM -I 100 -N -1 -S 100 -M 1.0E-6

Cluster mode

☒ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☐ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

Clusterer output

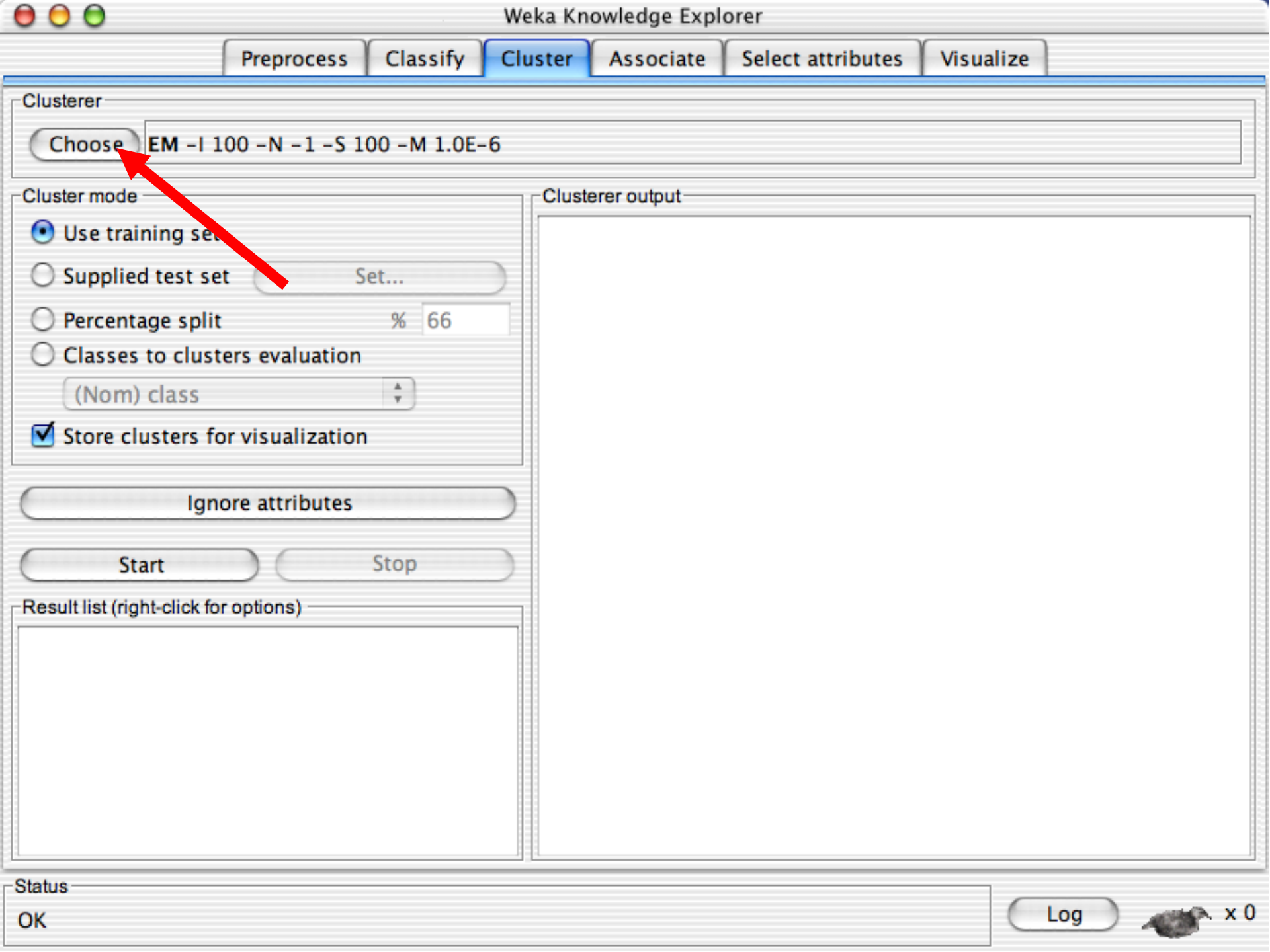
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

EM -I 100 -N -1 -S 100 -M 1.0E-6

Cluster mode

☒ Use training set☐ Supplied test set

Set...

☐ Percentage split

%

66

☐ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

Clusterer output

Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

- weka
  - clusters
    - EM
    - SimpleKMeans
    - Cobweb
    - FarthestFirst
    - XMeans

77387815

Clusterer output

Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☒ Use training set☐ Supplied test set

Set...

☐ Percentage split

%

66

☐ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

Clusterer output

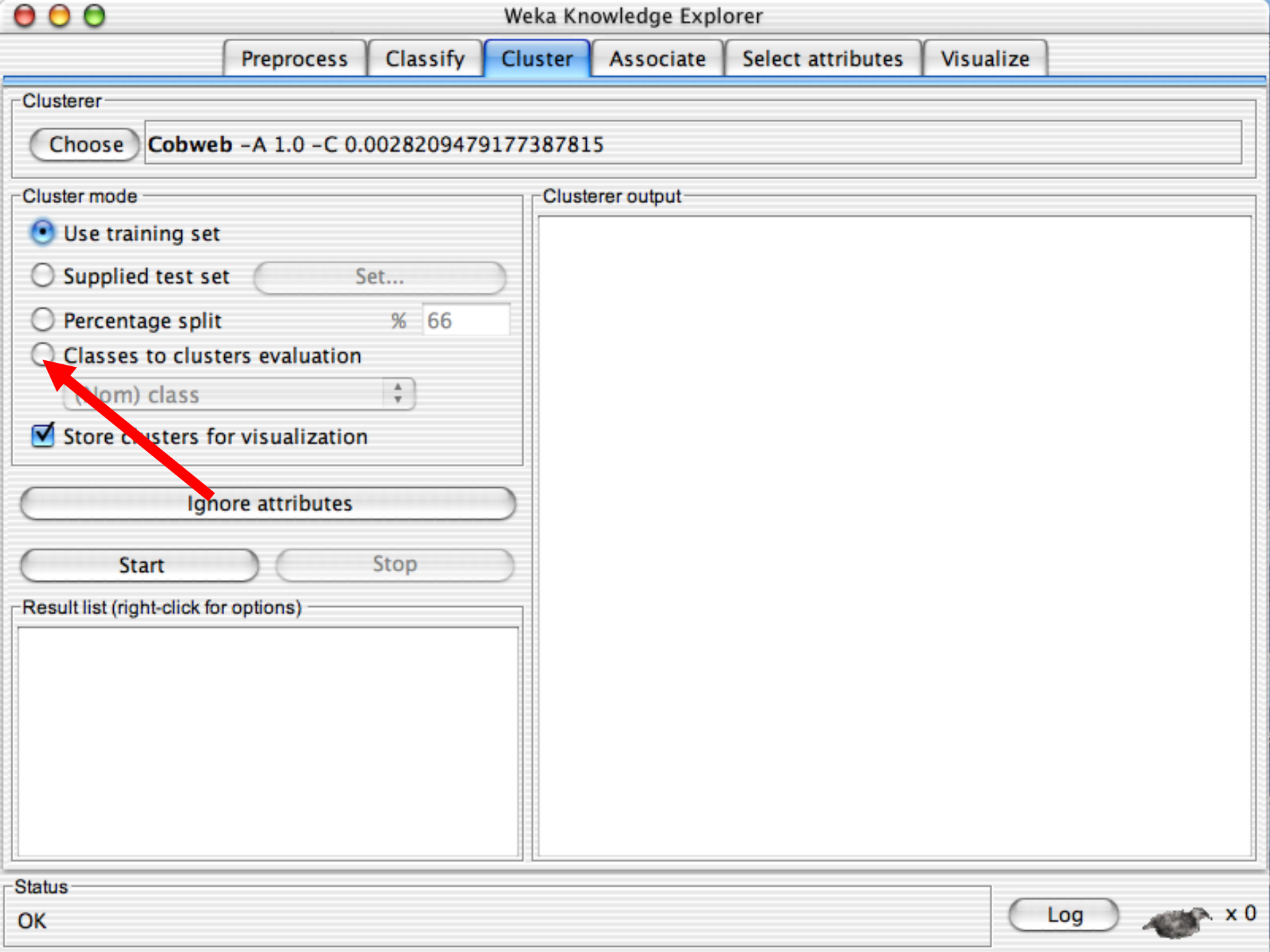
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☒ Use training set

☐ Supplied test set

Set...

☐ Percentage split

%

66

☐ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

Clusterer output

Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☐ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☒ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

Clusterer output

Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☐ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☒ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

Clusterer output

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☐ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☒ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

16:05:58 - Cobweb

Clusterer output

=== Run information ===

Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917  
Relation: iris  
Instances: 150  
Attributes: 5  
sepallength  
sepalwidth  
petallength  
petalwidth

Ignored:

class

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

Number of merges: 0  
Number of splits: 0  
Number of clusters: 3

node 0 [ 150]  
| leaf 1 [ 96]  
node 0 [ 150]  
| leaf 2 [ 54]

=== Evaluation on training set ===



Status

OK

Log

 x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☐ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☒ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

16:05:58 - Cobweb

Clusterer output

=== Run information ===

Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917  
Relation: iris  
Instances: 150  
Attributes: 5  
sepallength  
sepalwidth  
petallength  
petalwidth

Ignored:

class

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

Number of merges: 0  
Number of splits: 0  
Number of clusters: 3

node 0 [ 150]  
| leaf 1 [ 96]  
node 0 [ 150]  
| leaf 2 [ 54]

=== Evaluation on training set ===

Log

x 0

Status

OK



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☐ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☒ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

16:05:58 - Cobweb

Clusterer output

Number of clusters: 3

node 0 [ 150]

| leaf 1 [ 96]

node 0 [ 150]

| leaf 2 [ 54]

Clustered Instances

1 100 ( 67%)

2 50 ( 33%)

Class attribute: class

Classes to Clusters:

1 2 &lt;-- assigned to cluster

0 50 | Iris-setosa

50 0 | Iris-versicolor

50 0 | Iris-virginica

Cluster 1 &lt;-- Iris-versicolor

Cluster 2 &lt;-- Iris-setosa

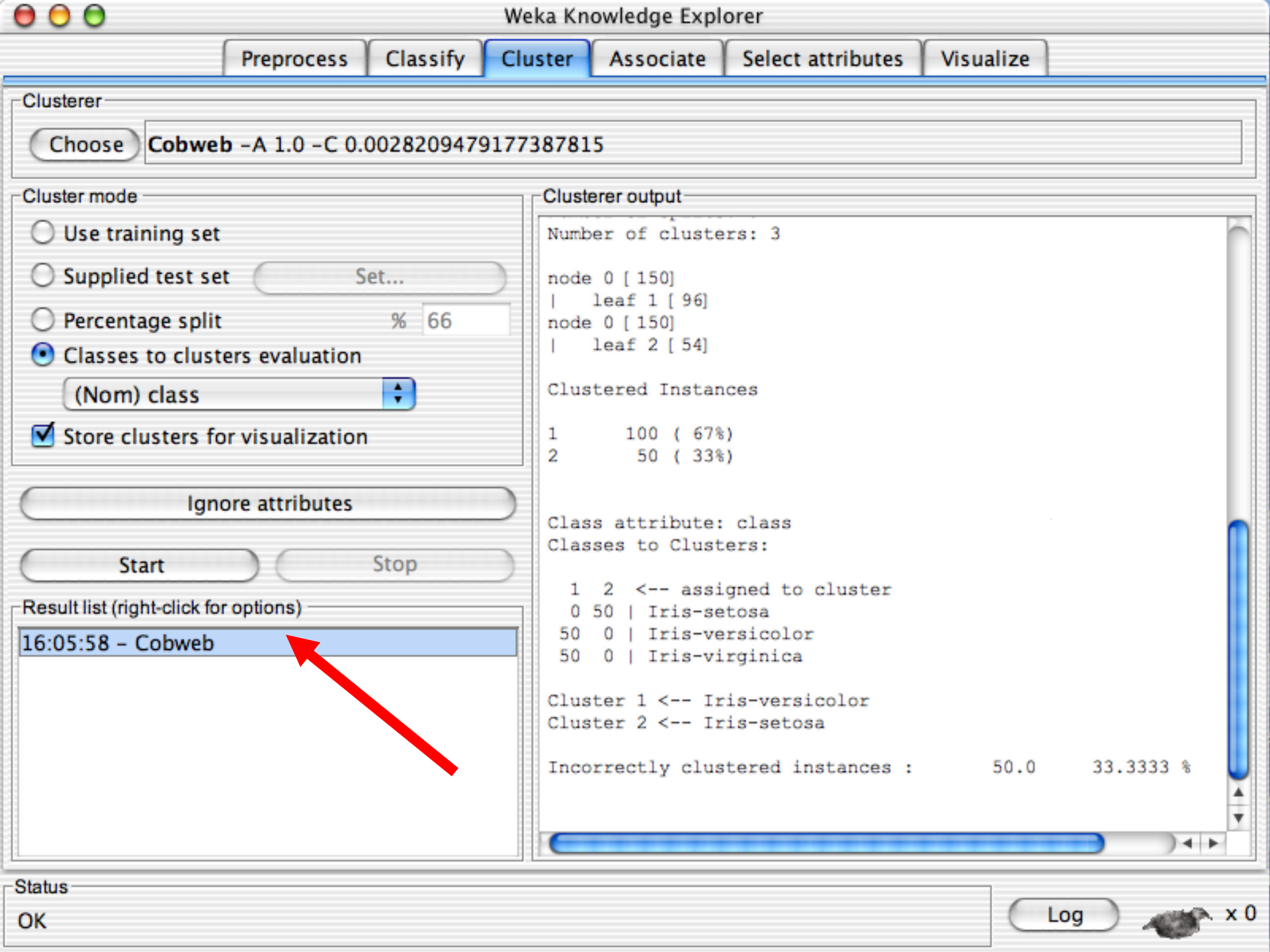
Incorrectly clustered instances : 50.0 33.3333 %

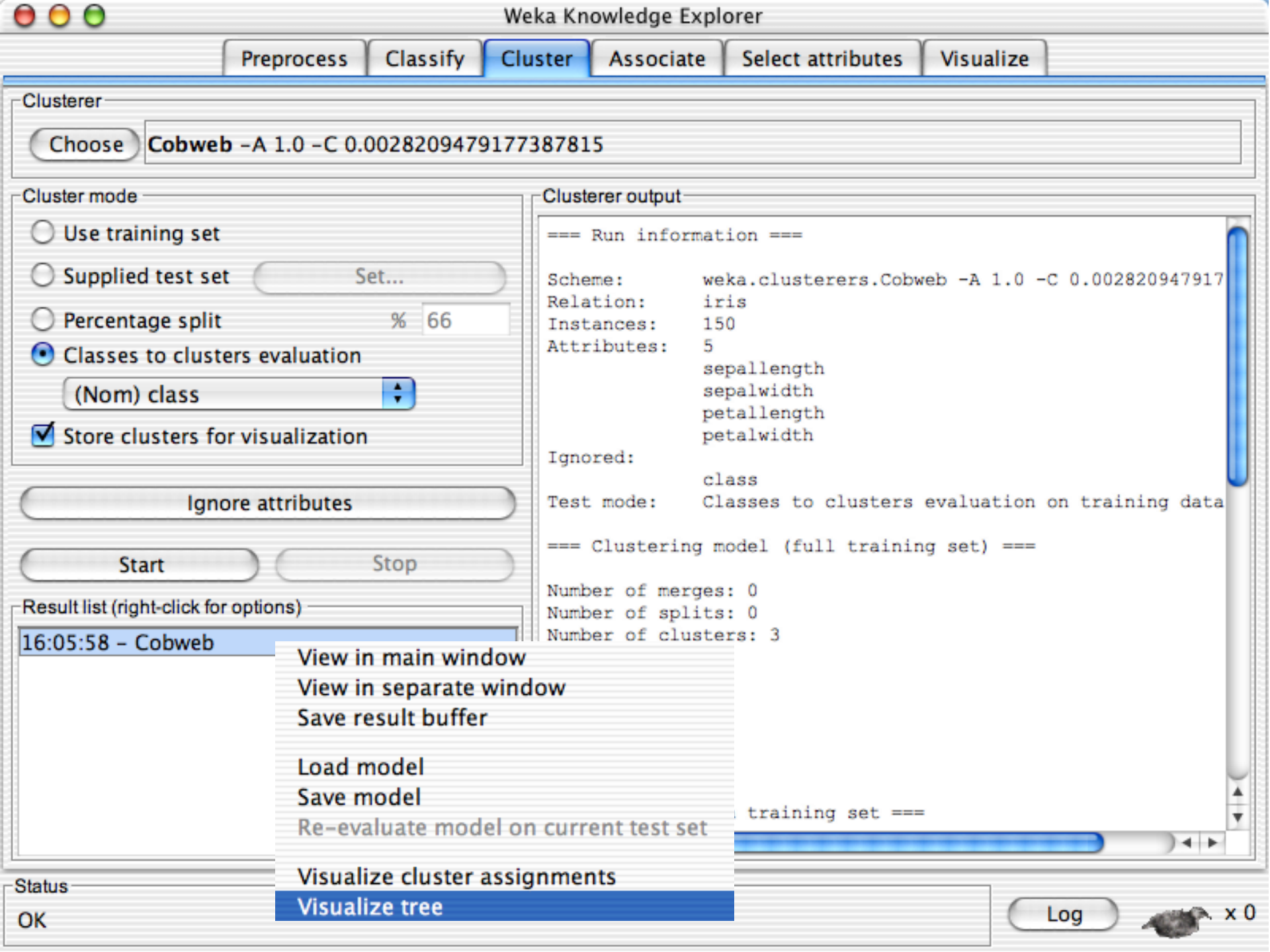
Status

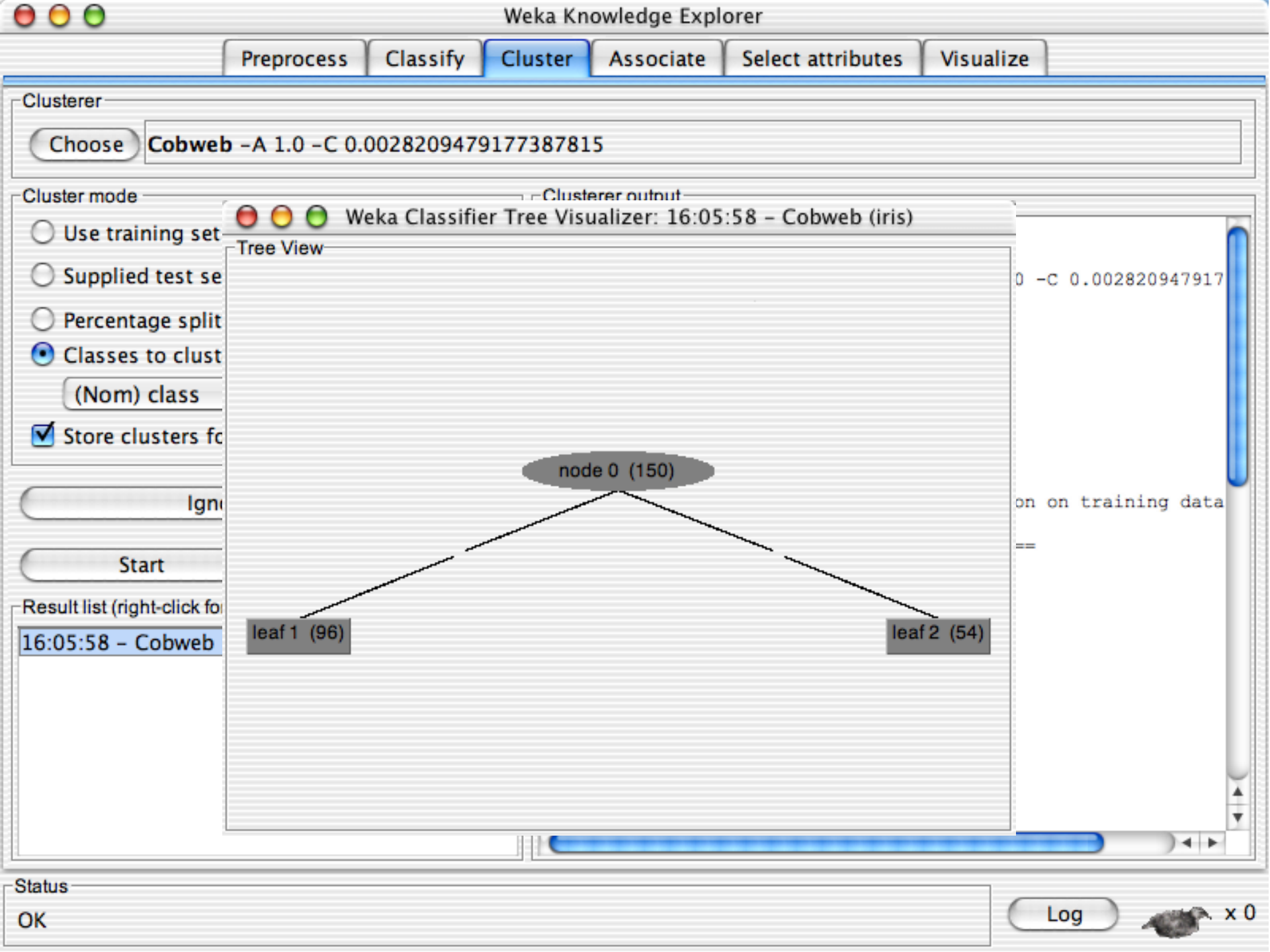
OK

Log

 x 0









Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

☐ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☒ Classes to clusters evaluation

(Nom) class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

16:05:58 - Cobweb

View in main window

View in separate window

Save result buffer

Load model

Save model

Re-evaluate model on current test set

Visualize cluster assignments

Visualize tree

Clusterer output

=== Run information ===

Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917  
Relation: iris  
Instances: 150  
Attributes: 5  
sepallength  
sepalwidth  
petallength  
petalwidth

Ignored:

class

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

Number of merges: 0  
Number of splits: 0  
Number of clusters: 3

on training set ===

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Clusterer

Choose

Cobweb -A 1.0 -C 0.0028209479177387815



Weka Clusterer Visualize: 16:05:58 - Cobweb (iris)

Cluster mode

☐ Use training se

X: petallength (Num)



Y: petalwidth (Num)

☐ Supplied test s

Colour: Cluster (Nom)



Select Instance

☐ Percentage spli☒ Classes to clus

Reset

Clear

Save

Jitter



(Nom) class

Plot: iris\_clustered

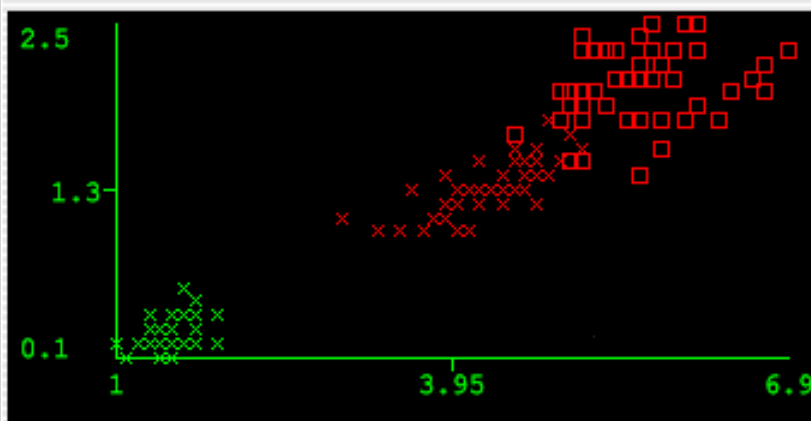
☒ Store clusters f

lgr

Start

Result list (right-click f

16:05:58 - Cobweb

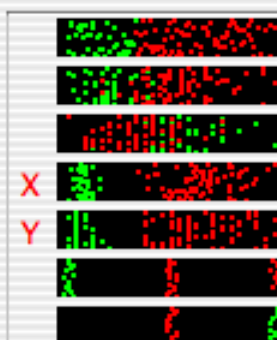


Class colour

cluster0

cluster1

cluster2



ion on training data

=== Evaluation on training set ===



Status

OK

Log



# Explorer: finding associations

- WEKA contains an implementation of the Apriori algorithm for learning association rules
  - ◆ Works only with discrete data
- Can identify statistical dependencies between groups of attributes:
  - ◆ milk, butter  $\Rightarrow$  bread, eggs (with confidence 0.9 and support 2000)
- Apriori can compute all rules that have a given minimum support and exceed a given confidence



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Associator

Choose

Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0

Start

Stop

Result list (right-click for options)

Associator output

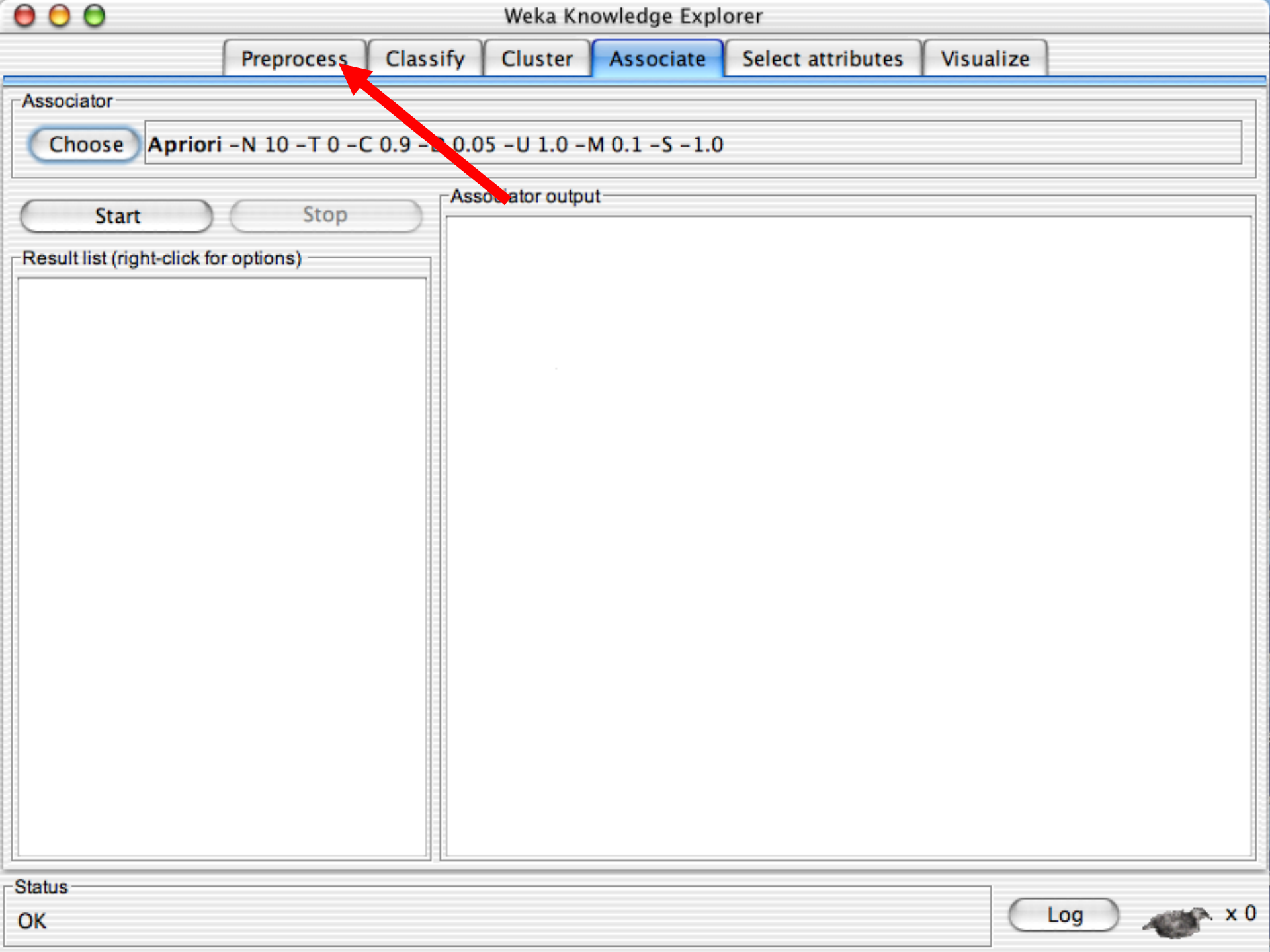
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Associator

Choose

Apriori -N 10 -T 0 -C 0.9 -L 0.05 -U 1.0 -M 0.1 -S -1.0

Start

Stop

Result list (right-click for options)

Associator output

Status

OK

Log

x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

None

Apply

Current relation

Relation: vote

Instances: 435

Attributes: 17

Attributes

| No. | Name                                   |
|-----|----------------------------------------|
| 1   | handicapped-infants                    |
| 2   | water-project-cost-sharing             |
| 3   | adoption-of-the-budget-resolution      |
| 4   | physician-fee-freeze                   |
| 5   | el-salvador-aid                        |
| 6   | religious-groups-in-schools            |
| 7   | anti-satellite-test-ban                |
| 8   | aid-to-nicaraguan-contras              |
| 9   | mx-missile                             |
| 10  | immigration                            |
| 11  | synfuels-corporation-cutback           |
| 12  | education-spending                     |
| 13  | superfund-right-to-sue                 |
| 14  | crime                                  |
| 15  | duty-free-exports                      |
| 16  | export-administration-act-south-africa |
| 17  | Class                                  |

Selected attribute

Name: handicapped-infants

Type: Nominal

Missing: 12 (3%)

Distinct: 2

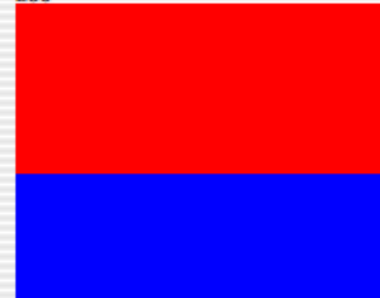
Unique: 0 (0%)

| Label | Count |
|-------|-------|
| n     | 236   |
| y     | 187   |

Colour: Class (Nom)

Visualize All

236



187



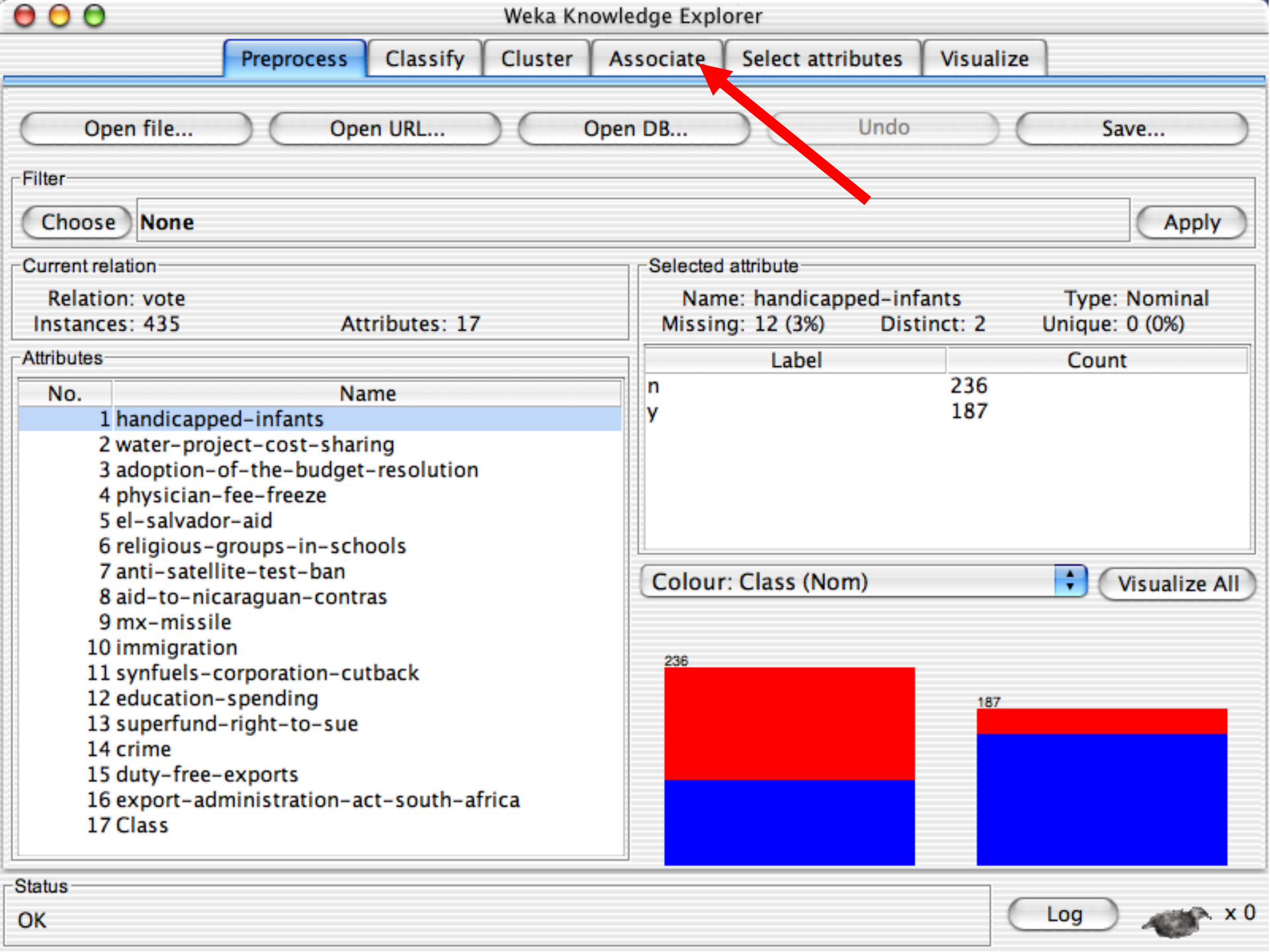
Status

OK

Log



x 0





Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Associator

Choose

Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0

Start

Stop

Result list (right-click for options)

Associator output

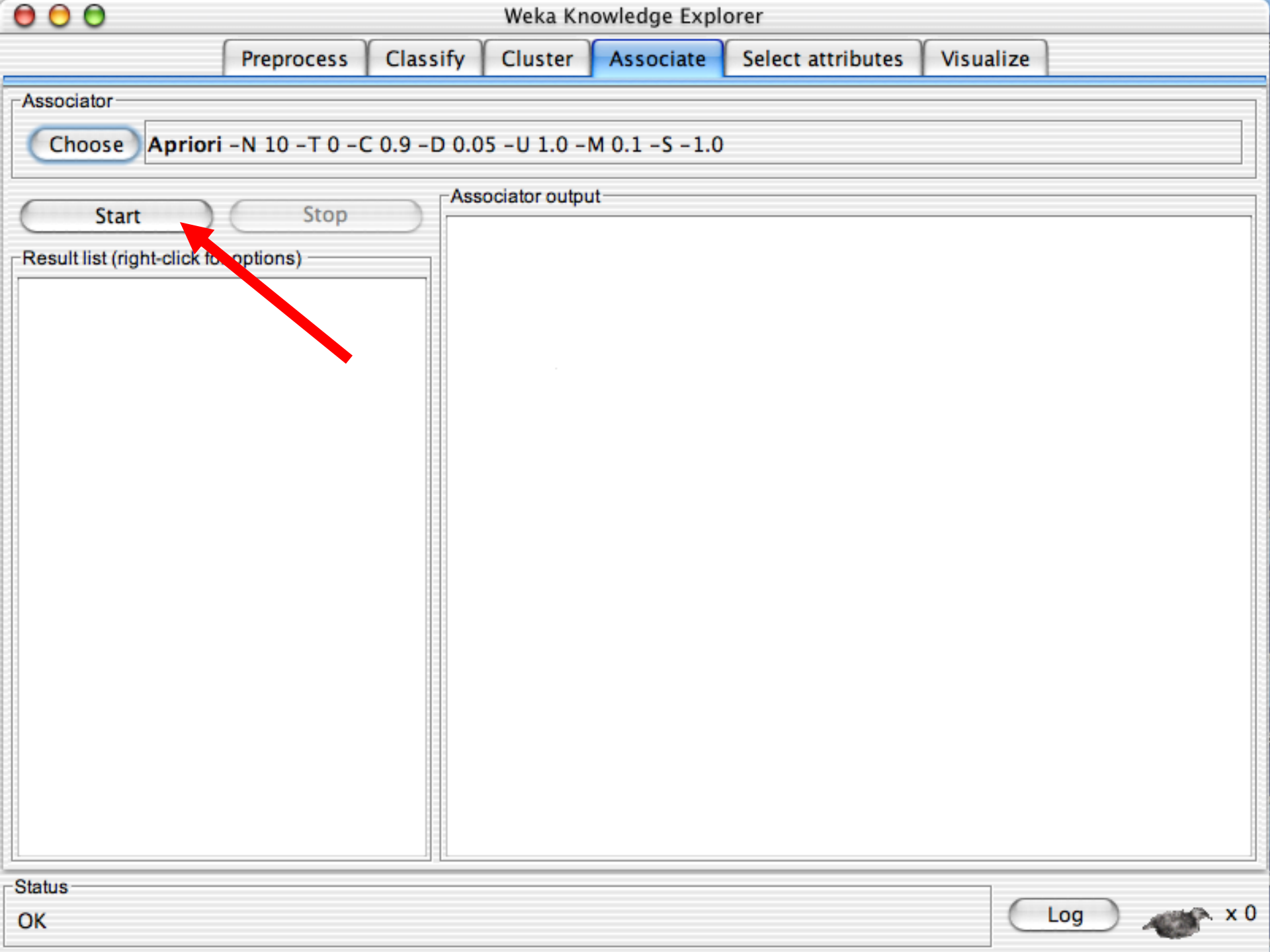
Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Associator

Choose

Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0

Start

Stop

Associator output

Result list (right-click for options)

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Associator

**Choose** Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**Start****Stop**

Result list (right-click for options)

16:29:37 - Apriori

Associator output

Minimum metric &lt;confidence&gt;: 0.9

Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20

Size of set of large itemsets L(2): 17

Size of set of large itemsets L(3): 6

Size of set of large itemsets L(4): 1

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 210 ==> Class=democrat 210
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 conf: (0.99)
5. physician-fee-freeze=n 247 ==> Class=democrat 245 conf: (0.99)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 conf: (0.98)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 conf: (0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 204
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 conf: (0.98)
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210

Status

OK

Log

 x 0

# Explorer: attribute selection

- Panel that can be used to investigate which (subsets of) attributes are the most predictive ones
- Attribute selection methods contain two parts:
  - ◆ A search method: best-first, forward selection, random, exhaustive, genetic algorithm, ranking
  - ◆ An evaluation method: correlation-based, wrapper, information gain, chi-squared, ...
- Very flexible: WEKA allows (almost) arbitrary combinations of these two



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Attribute Evaluator

Choose

CfsSubsetEval

Search Method

Choose

BestFirst -D 1 -N 5

Attribute Selection Mode



Use full training set



Cross-validation

Folds

10

Seed

1

(Nom) Class

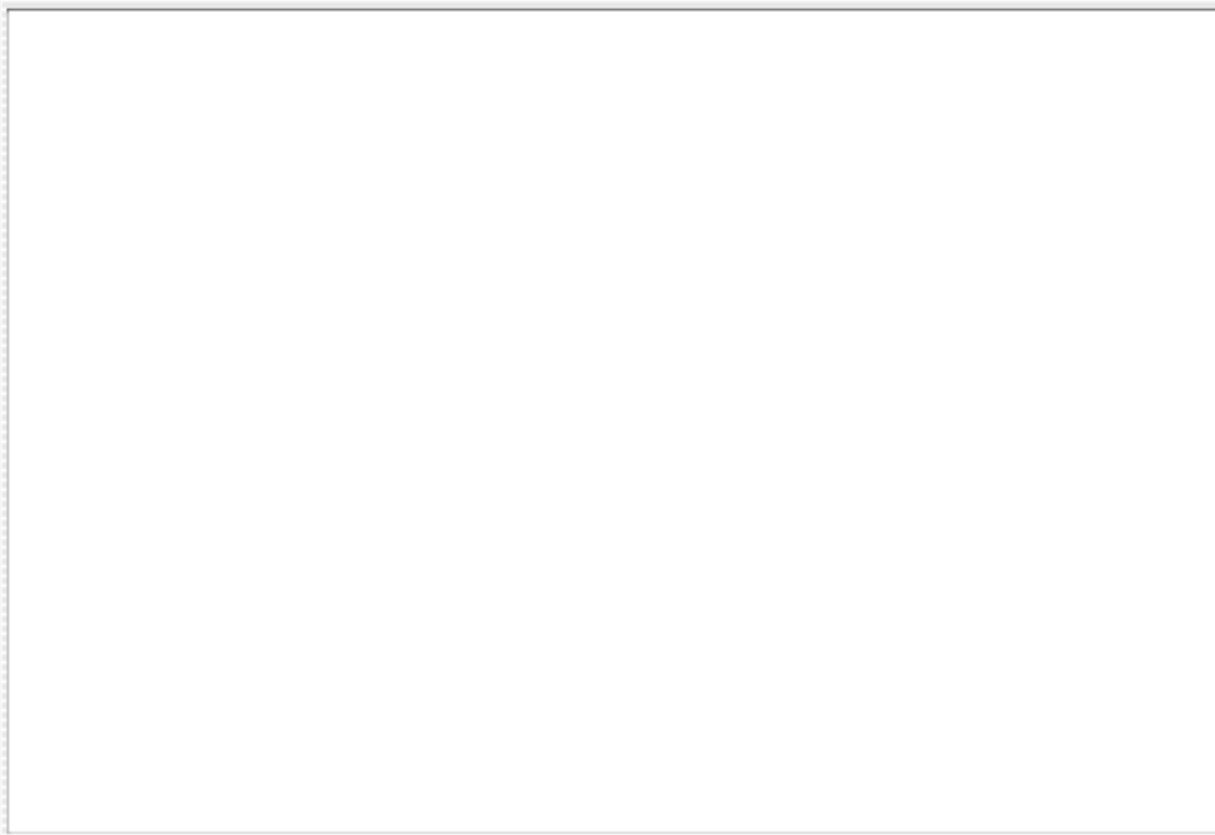


Start

Stop

Result list (right-click for options)

Attribute selection output



Status

OK

Log



x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Attribute Evaluator

Choose

CfsSubsetEval

Search Method

Choose

BestFirst -D 1 -N 5

Attribute Selection Mode



Use full training set



Cross-validation

Folds

10

Seed

1

(Nom) Class

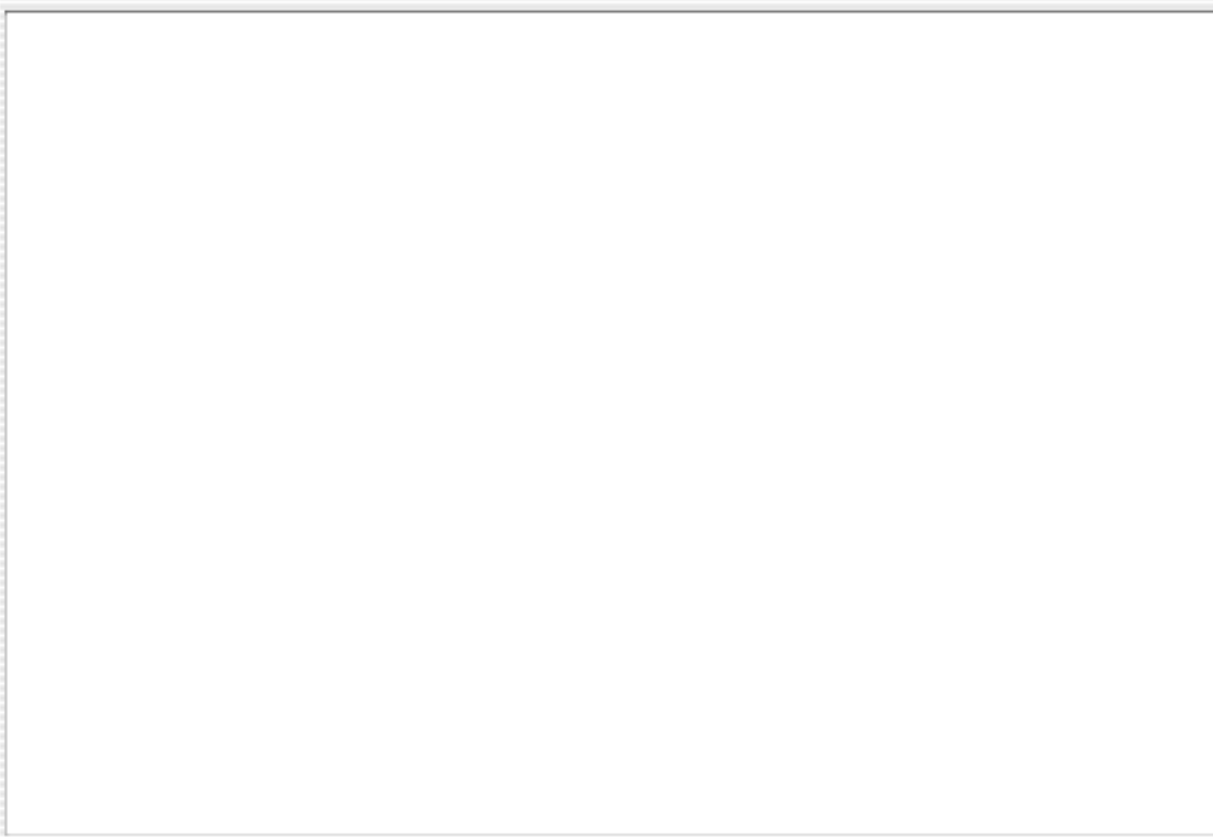


Start

Stop

Result list (right-click for options)

Attribute selection output



Status

OK

Log



x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Attribute Evaluator

Choose

CfsSubsetEval

## Search Method

Choose

BestFirst -D 1 -N 5

## Attribute Selection Mode



Use full training set



Cross-validation

Folds

10

Seed

1

(Nom) Class



Start

Stop

## Result list (right-click for options)

16:39:40 - BestFirst + CfsSubsetEval

## Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
```

```
Evaluation mode:    evaluate on all training data
```

```
=== Attribute Selection on all input data ===
```

## Search Method:

```
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 83
Merit of best subset found:    0.729
```

```
Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
```

```
Selected attributes: 4 : 1
                    physician-fee-freeze
```

Status

OK

Log

x 0

Preprocess

Classify

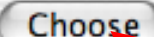
Cluster

Associate

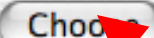
Select attributes

Visualize

## Attribute Evaluator

 **CfsSubsetEval**

## Search Method

 **BestFirst -D 1 -N 5**

## Attribute Selection Mode

☒ Use full training set☐ Cross-validation

Folds 10

Seed 1

(Nom) Class

Start

Stop

## Result list (right-click for options)

16:39:40 - BestFirst + CfsSubsetEval

## Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 83
  Merit of best subset found:    0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
  CFS Subset Evaluator

Selected attributes: 4 : 1
                    physician-fee-freeze
```

Status

OK

Log

 x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Attribute Evaluator

- weka
  - attributeSelection
    - CfsSubsetEval
    - ClassifierSubsetEval
    - WrapperSubsetEval
    - ConsistencySubsetEval
    - ReliefFAttributeEval
    - InfoGainAttributeEval
    - GainRatioAttributeEval
    - SymmetricalUncertAttributeEval
    - OneRAttributeEval
    - ChiSquaredAttributeEval
    - PrincipalComponents
    - SVMAttributeEval

## Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode:    evaluate on all training data

Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 83
  Merit of best subset found:    0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
  CFS Subset Evaluator

Selected attributes: 4 : 1
                    physician-fee-freeze
```

Status

OK

Log

 x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Attribute Evaluator

Choose

InfoGainAttributeEval

## Search Method

- weka
  - attributeSelection
    - BestFirst
    - ForwardSelection
    - RaceSearch
    - GeneticSearch
    - RandomSearch
    - ExhaustiveSearch
    - Ranker
    - RankSearch

E308 -N -1

## Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
```

```
evaluation mode:    evaluate on all training data
```

```
Attribute Selection on all input data ===
```

## Search Method:

```
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 83
Merit of best subset found:    0.729
```

```
Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
```

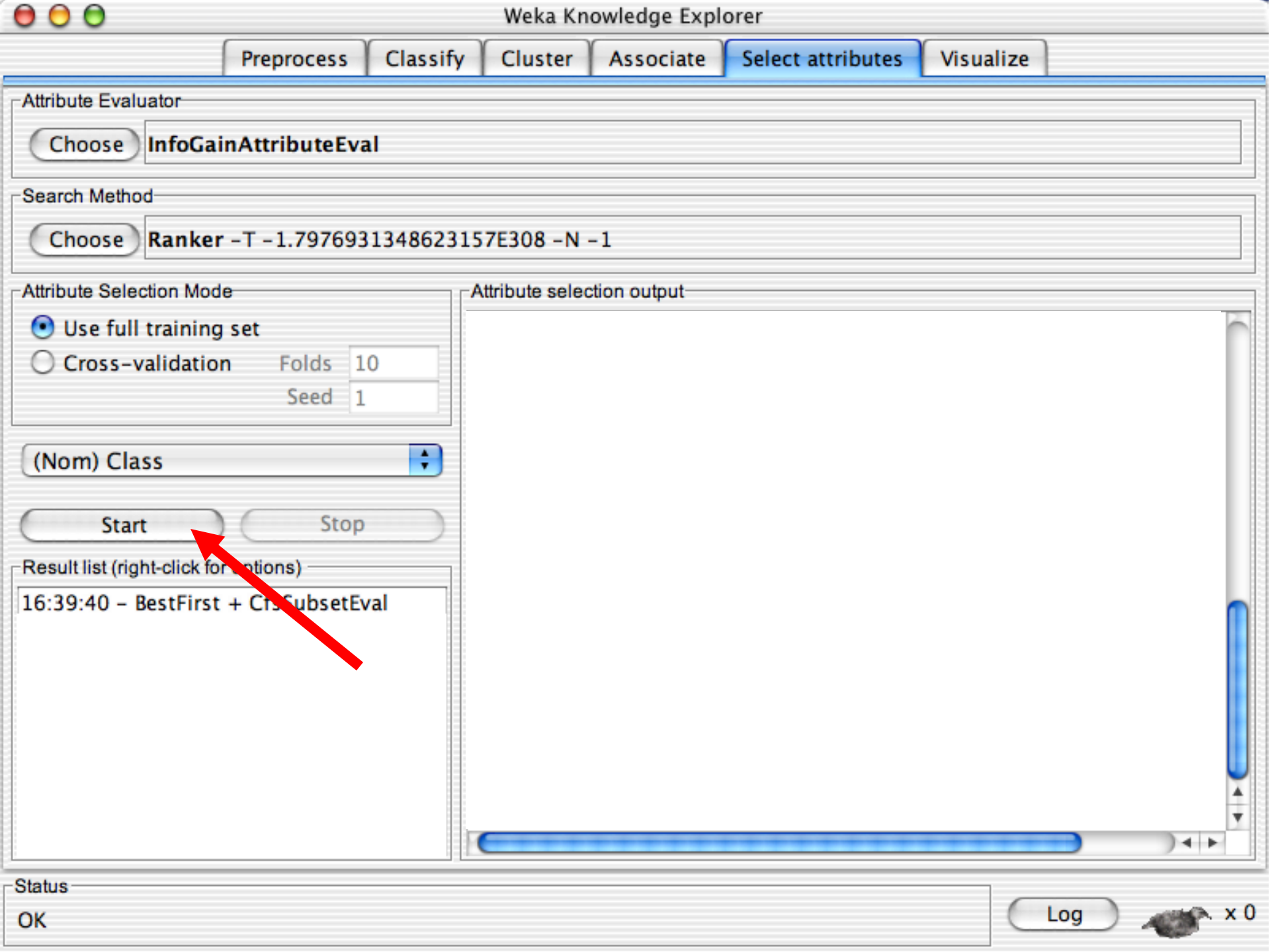
```
Selected attributes: 4 : 1
                    physician-fee-freeze
```

Status

OK

Log

 x 0



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Attribute Evaluator

Choose

InfoGainAttributeEval

## Search Method

Choose

Ranker -T -1.7976931348623157E308 -N -1

## Attribute Selection Mode



Use full training set



Cross-validation

Folds

10

Seed

1

(Nom) Class



Start

Stop

## Result list (right-click for options)

16:39:40 - BestFirst + CfsSubsetEval

16:43:05 - Ranker + InfoGainAttributeEval

## Attribute selection output

Information Gain Ranking Filter

## Ranked attributes:

|           |    |                                        |
|-----------|----|----------------------------------------|
| 0.7078541 | 4  | physician-fee-freeze                   |
| 0.4185726 | 3  | adoption-of-the-budget-resolution      |
| 0.4028397 | 5  | el-salvador-aid                        |
| 0.34036   | 12 | education-spending                     |
| 0.3123121 | 14 | crime                                  |
| 0.3095576 | 8  | aid-to-nicaraguan-contras              |
| 0.2856444 | 9  | mx-missile                             |
| 0.2121705 | 13 | superfund-right-to-sue                 |
| 0.2013666 | 15 | duty-free-exports                      |
| 0.1902427 | 7  | anti-satellite-test-ban                |
| 0.1404643 | 6  | religious-groups-in-schools            |
| 0.1211834 | 1  | handicapped-infants                    |
| 0.1007458 | 11 | synfuels-corporation-cutback           |
| 0.0529956 | 16 | export-administration-act-south-africa |
| 0.0049097 | 10 | immigration                            |
| 0.0000117 | 2  | water-project-cost-sharing             |

Selected attributes: 4,3,5,12,14,8,9,13,15,7,6,1,11,16,10,2 : 16



Status

OK

Log

x 0

# Explorer: data visualization

- Visualization very useful in practice: e.g. helps to determine difficulty of the learning problem
- WEKA can visualize single attributes (1-d) and pairs of attributes (2-d)
  - ◆ To do: rotating 3-d visualizations (Xgobi-style)
- Color-coded class values
- “Jitter” option to deal with nominal attributes (and to detect “hidden” data points)
- “Zoom-in” function



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

Choose

None

Apply

Current relation

Relation: Glass

Instances: 214

Attributes: 10

Attributes

| No. | Name |
|-----|------|
| 1   | RI   |
| 2   | Na   |
| 3   | Mg   |
| 4   | Al   |
| 5   | Si   |
| 6   | K    |
| 7   | Ca   |
| 8   | Ba   |
| 9   | Fe   |
| 10  | Type |

Selected attribute

Name: RI

Missing: 0 (0%)

Distinct: 178

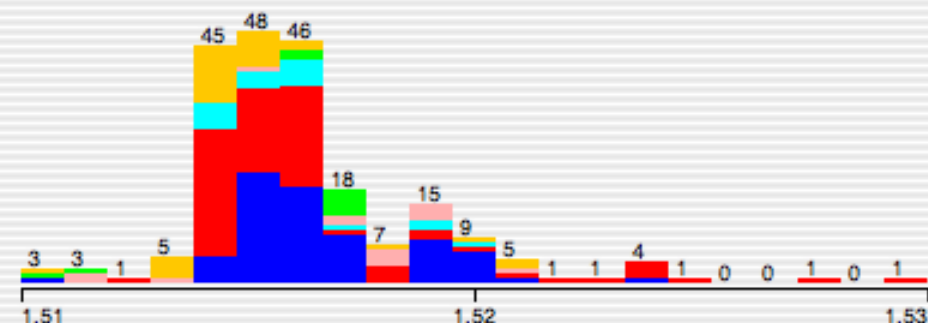
Type: Numeric

Unique: 145 (68%)

| Statistic | Value |
|-----------|-------|
| Minimum   | 1.511 |
| Maximum   | 1.534 |
| Mean      | 1.518 |
| StdDev    | 0.003 |

Colour: Type (Nom)

Visualize All



Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Plot Matrix

Ri

Na

Mg

Al

Si

K

Type

Fe

PlotSize: [100]

PointSize: [1]

Jitter:

Colour: Type (Nom)

Update

Select Attributes

SubSample % :

100

Class Colour

```
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps
```

Status

OK

Log

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Plot Matrix

Ri

Na

Mg

Al

Si

K

Type

Fe

PlotSize: [100]

PointSize: [1]

Jitter:

Colour: Type (Nom)

Update

Select Attributes

SubSample % : 100

Class Colour

```
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps
```

Status

OK

Log

x 0



Preprocess

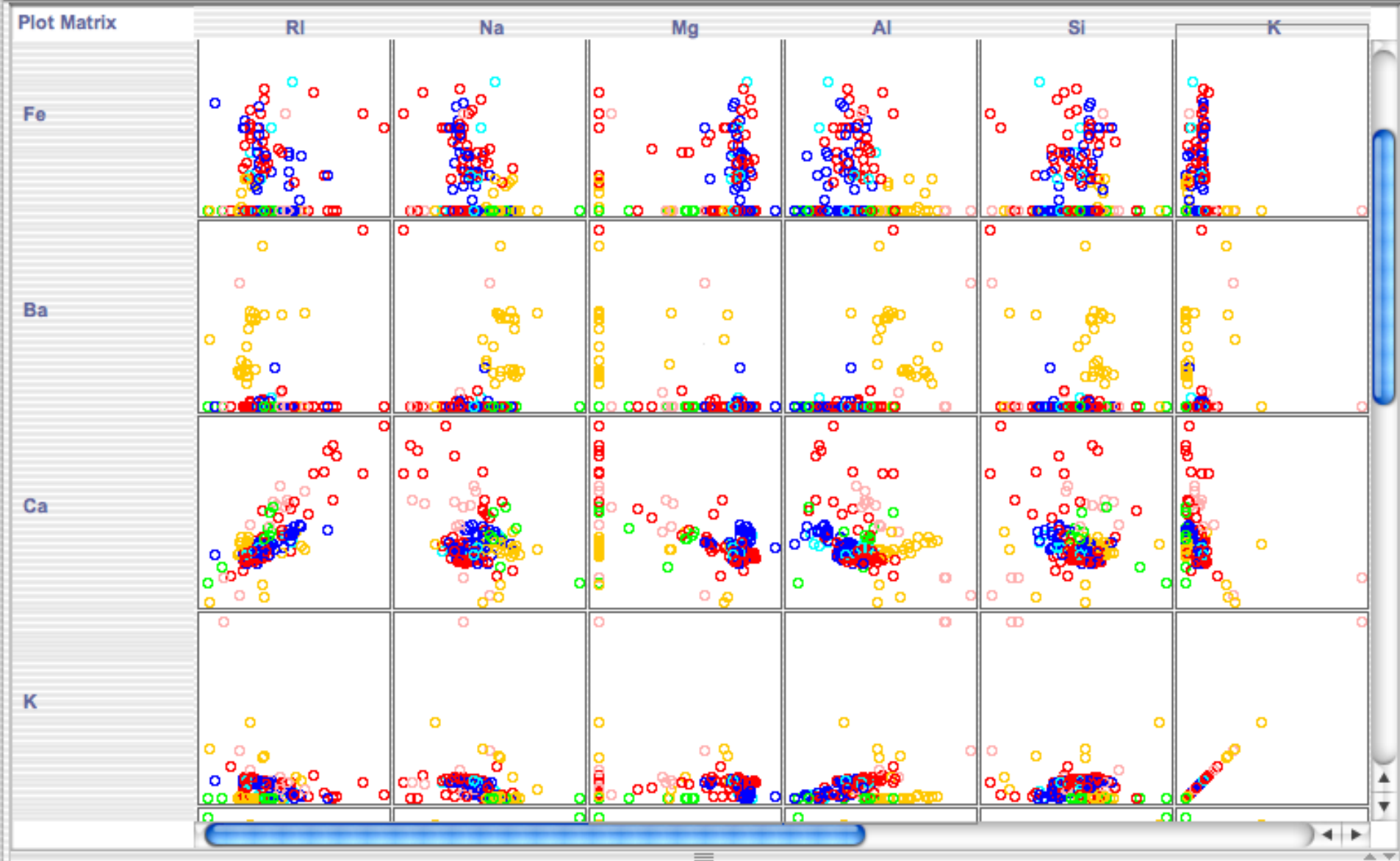
Classify

Cluster

Associate

Select attributes

Visualize



Status

OK

Log

x 0



Preprocess

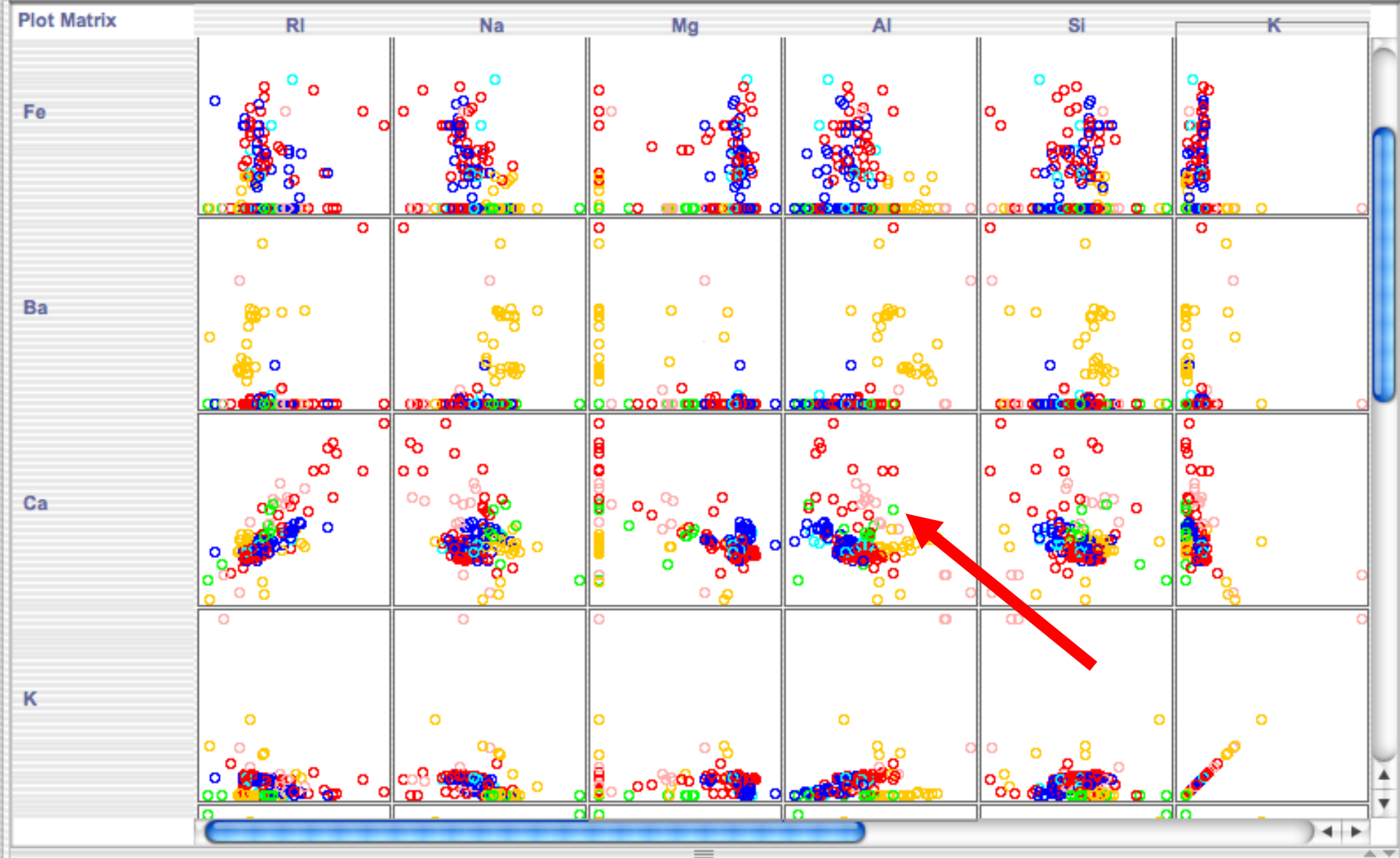
Classify

Cluster

Associate

Select attributes

Visualize



Status

OK

Log

x 0

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Select Instance

Reset

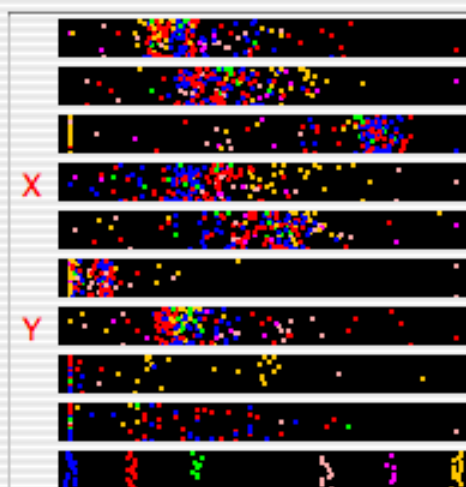
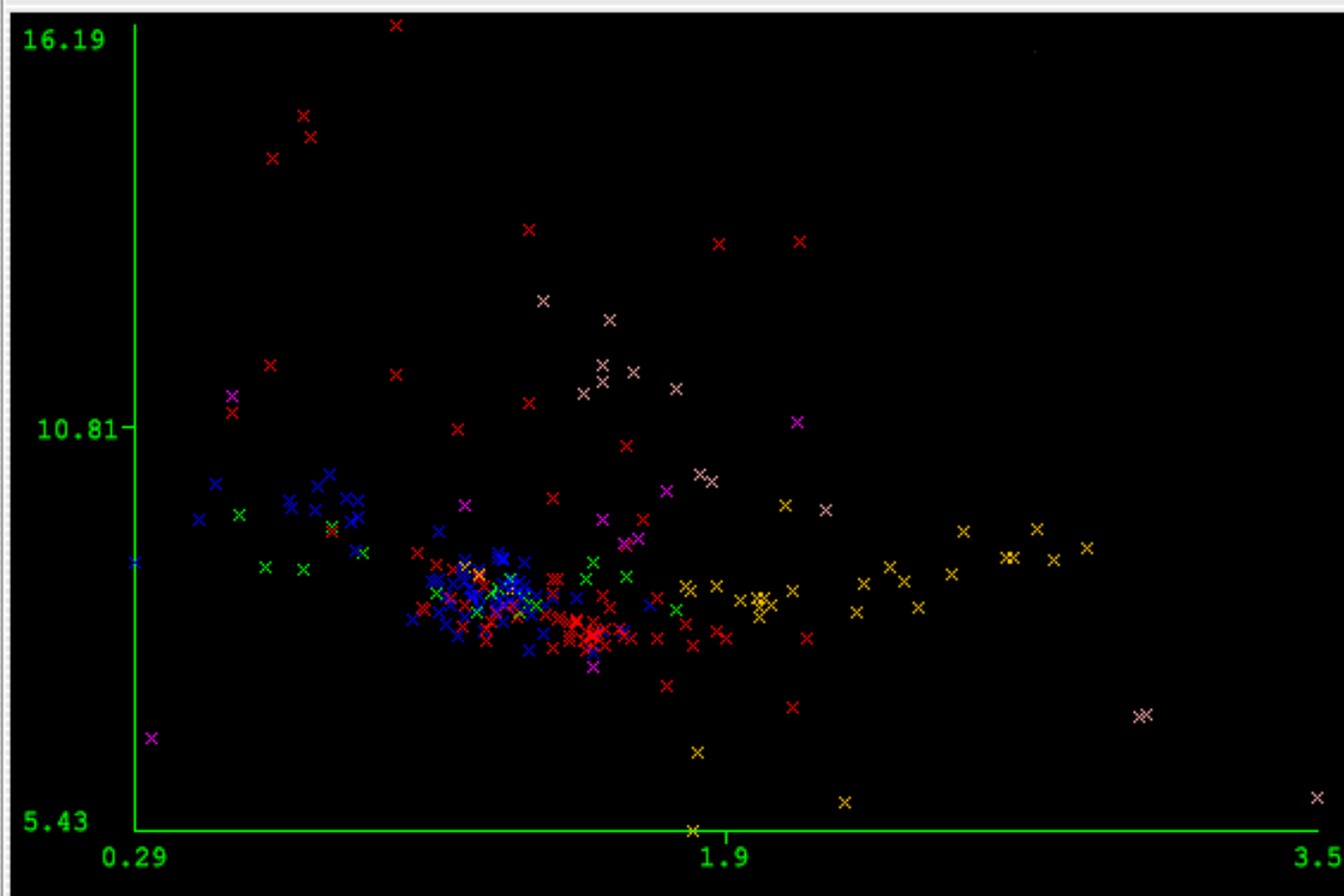
Clear

Save

Jitter



Plot: Glass



Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Select Instance

Reset

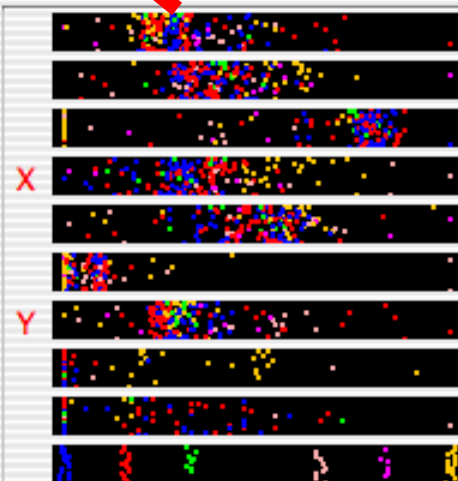
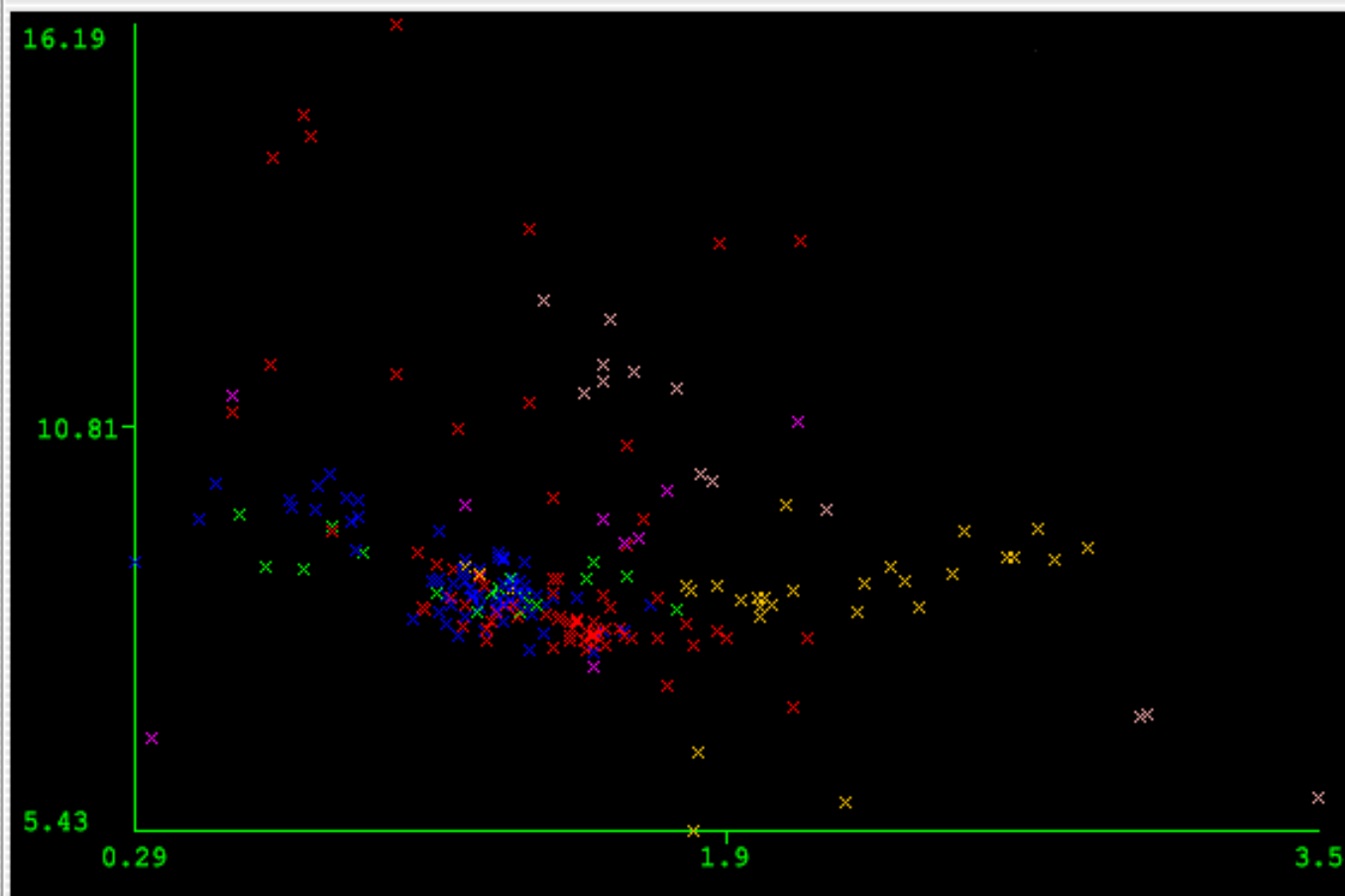
Clear

Save

Jitter



Plot: Glass



Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Rectangle

Submit

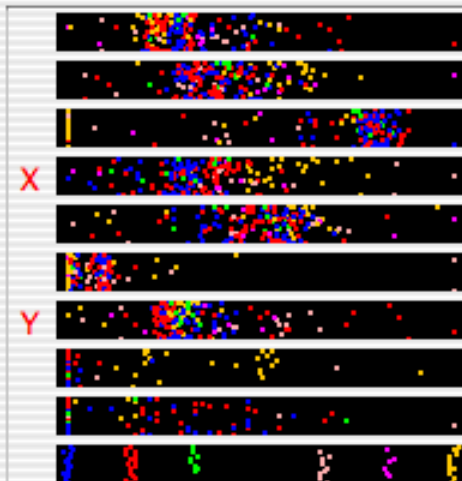
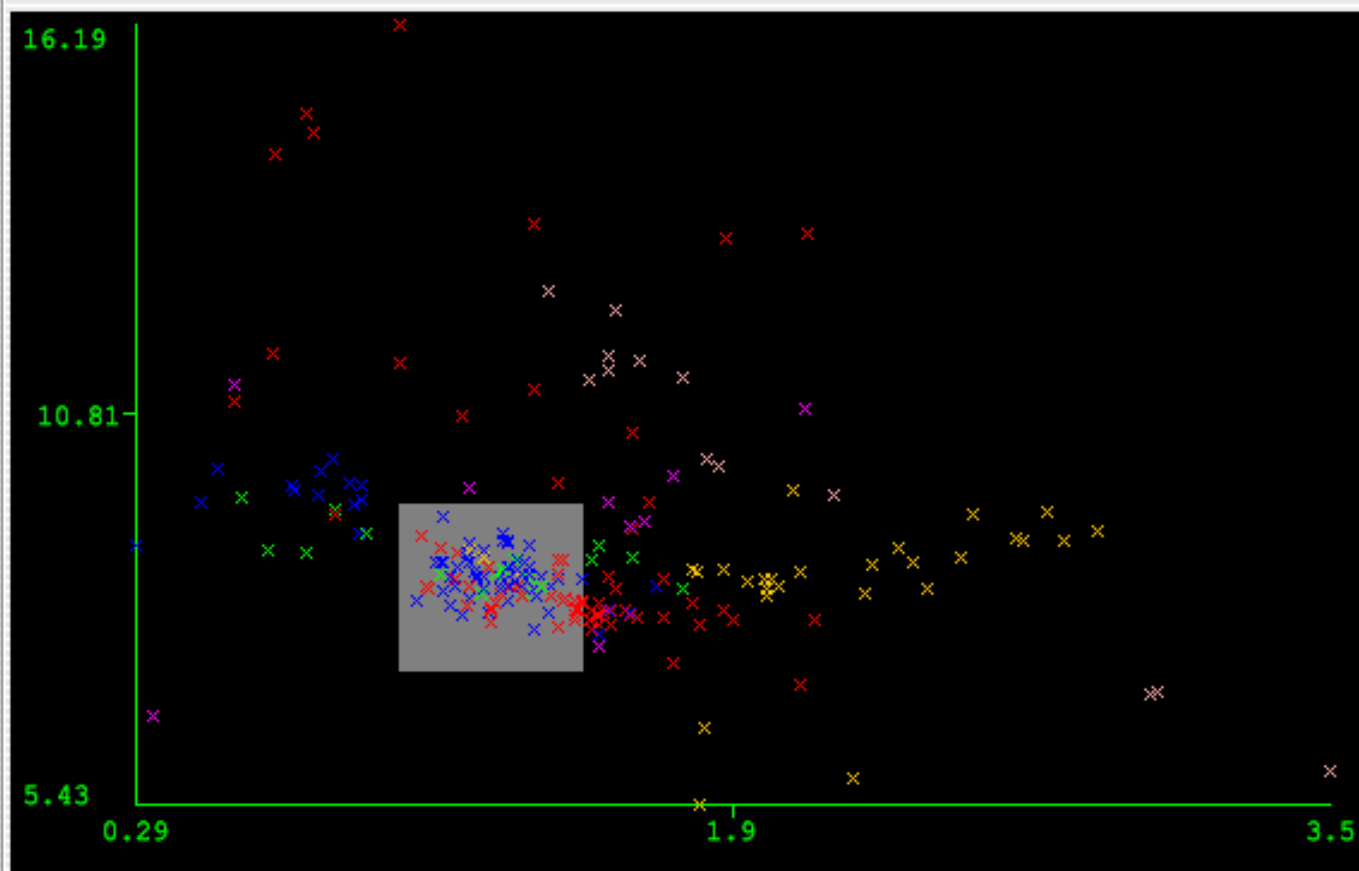
Clear

Save

Jitter



Plot: Glass



Class colour

build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Rectangle

Submit

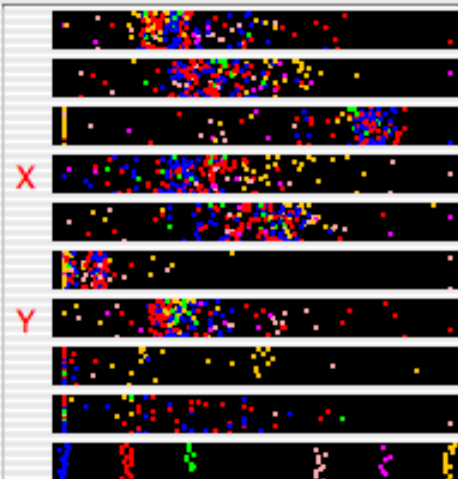
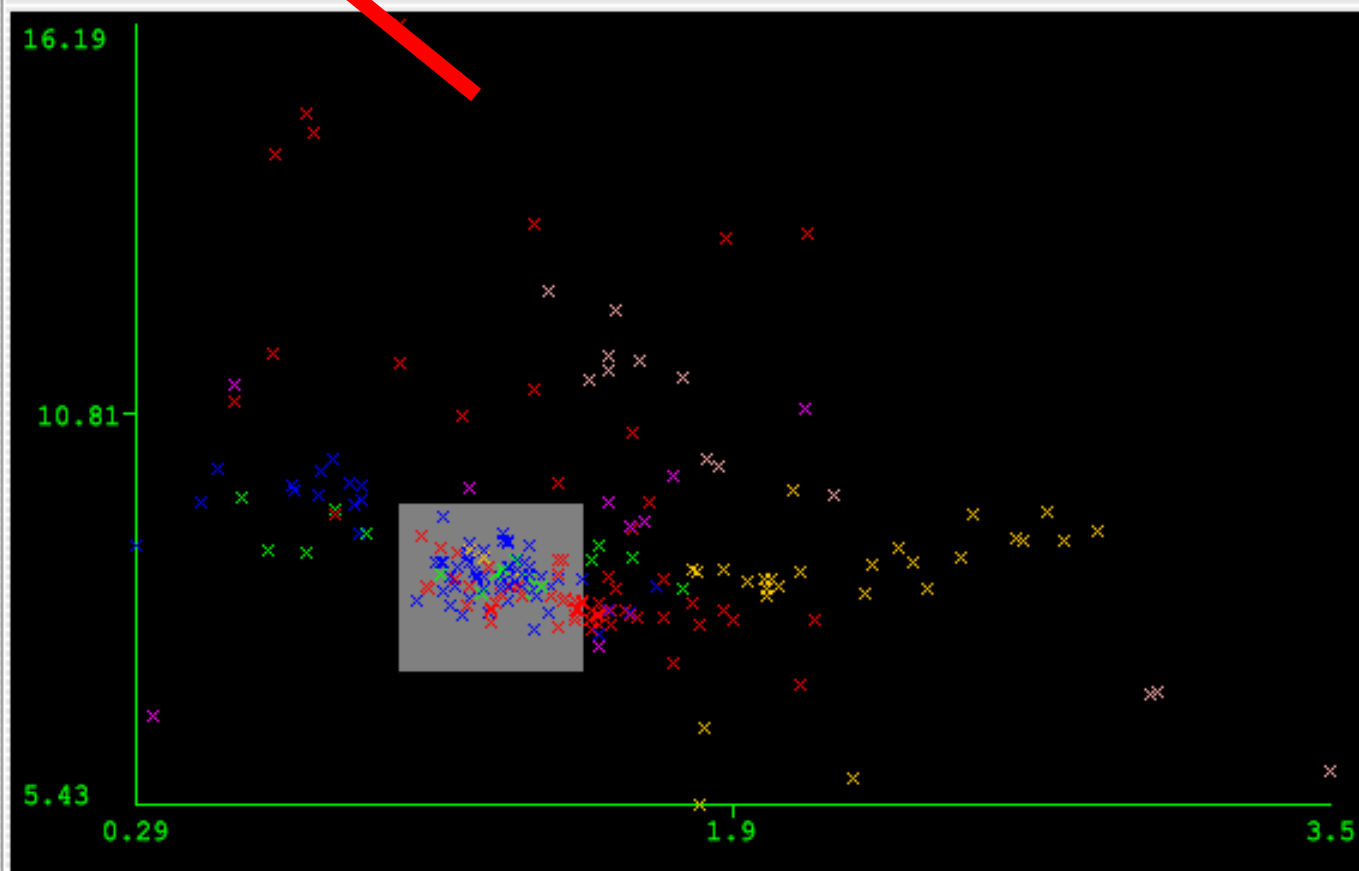
Clear

Save

Jitter



Plot: Glass



Class colour

build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Rectangle

Reset

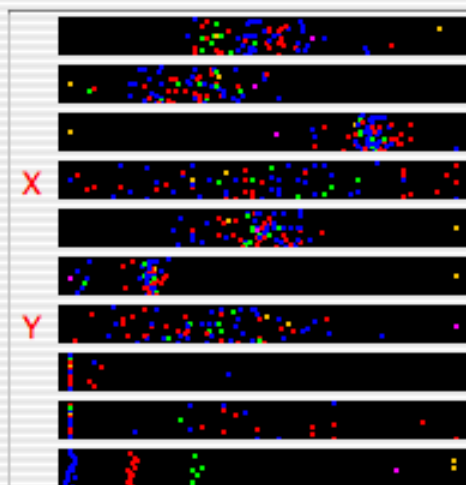
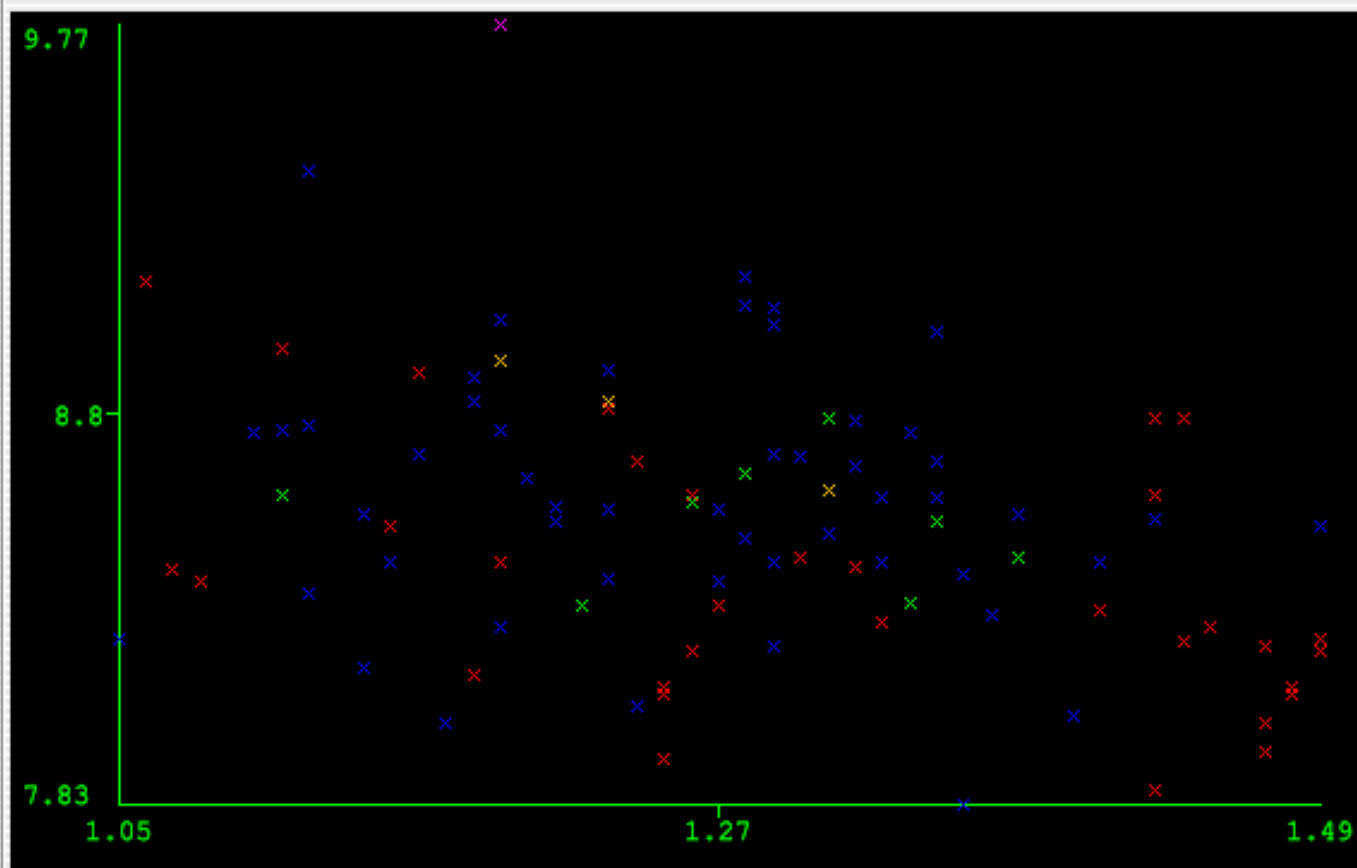
Clear

Save

Jitter



Plot: Glass



Class colour

build wind float

vehic wind non-float

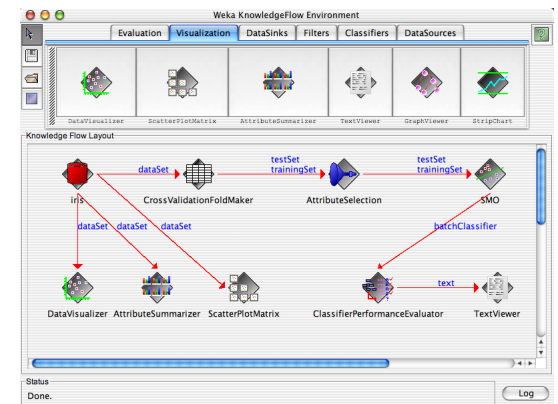
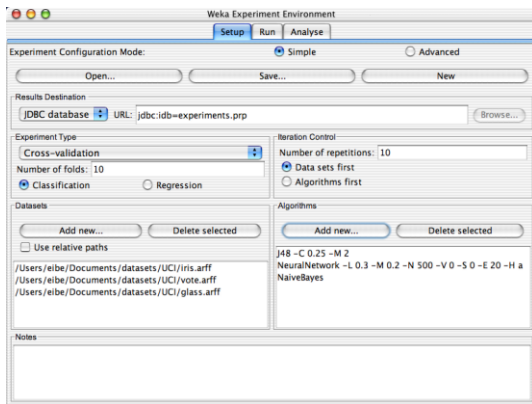
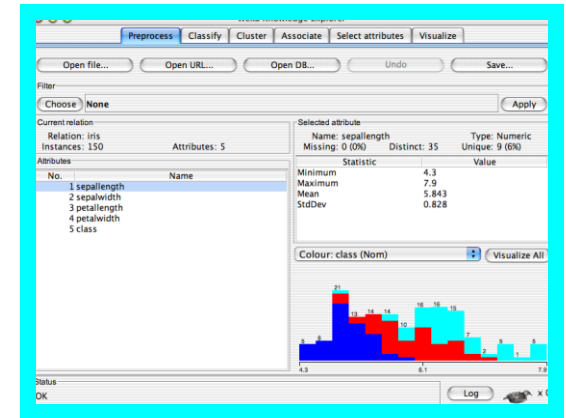
build wind non-float

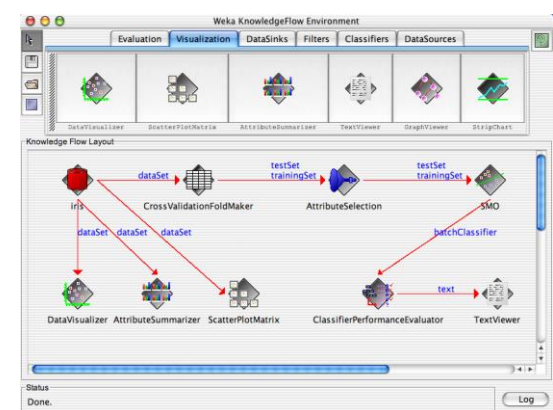
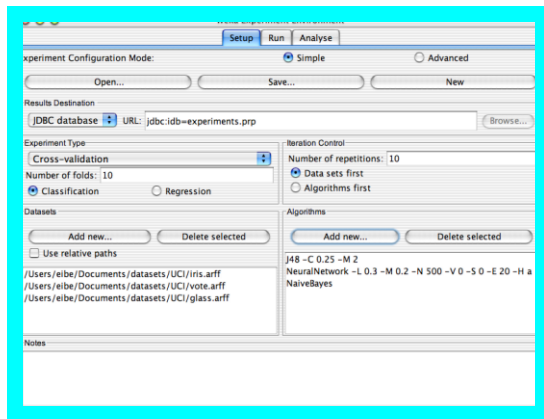
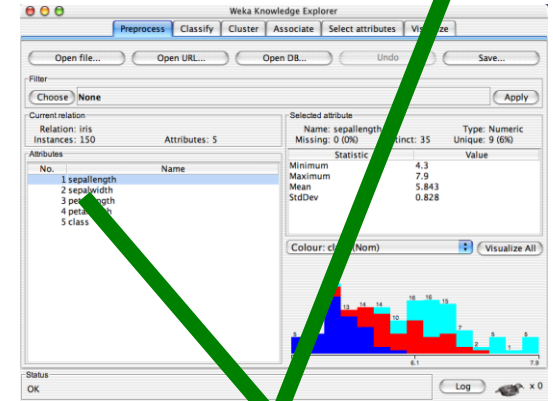
containers

tableware

vehic wind float

headlamps





# Performing experiments

- Experimenter makes it easy to compare the performance of different learning schemes
- For classification and regression problems
- Results can be written into file or database
- Evaluation options: cross-validation, learning curve, hold-out
- Can also iterate over different parameter settings
- Significance-testing built in!



Setup

Run

Analyse

Experiment Configuration Mode:

☒ Simple☐ Advanced

Open...

Save...

New

Results Destination

JDBC database ▴ ▾

Filename:

Browse...

Experiment Type

Cross-validation ▴ ▾

Number of folds:

☒ Classification☐ Regression

Iteration Control

Number of repetitions:

☒ Data sets first☐ Algorithms first

Datasets

Add new...

Delete selected

☐ Use relative paths

Algorithms

Add new...

Delete selected

Notes



Setup

Run

Analyse

Experiment Configuration Mode:

☒ Simple☐ Advanced

Open...

Save...

New

Results Destination

JDBC database ▾

Filename:

Browse...

Experiment Type

Cross-validation ▾

Number of folds:

☒ Classification☐ Regression

Iteration Control

Number of repetitions:

☒ Data sets first☐ Algorithms first

Datasets

Add new...

Delete selected

☐ Use relative paths

Algorithms

Add new...

Delete selected

Notes

Setup

Run

Analyse

Experiment Configuration Mode:

☒ Simple☐ Advanced

Open...

Save...

New

Results Destination

JDBC database

URL: jdbc:ide=experiments.prp

Browse...

Experiment Type

Cross-validation

Number of folds: 10

☒ Classification☐ Regression

Iteration Control

Number of repetitions: 10

☒ Data sets first☐ Algorithms first

Datasets

Add new...

Delete selected

☐ Use relative paths

/Users/eibe/Documents/datasets/UCI/iris.arff  
/Users/eibe/Documents/datasets/UCI/vote.arff  
/Users/eibe/Documents/datasets/UCI/glass.arff

Algorithms

Add new...

Delete selected

J48 -C 0.25 -M 2  
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a  
NaiveBayes

Notes

## Weka Experiment Environment

Setup

Run

Analyse

Experiment Configuration Mode:

☒ Simple☐ Advanced

Open...

Save...

New

Results Destination

JDBC database

URL: jdbc:ide=experiments.prp

Browse...

Experiment Type

Cross-validation

Number of folds: 10

☒ Classification☐ Regression

Iteration Control

Number of repetitions: 10

☒ Data sets first☐ Algorithms first

Datasets

Add new...

Delete selected

☐ Use relative paths

/Users/eibe/Documents/datasets/UCI/iris.arff  
/Users/eibe/Documents/datasets/UCI/vote.arff  
/Users/eibe/Documents/datasets/UCI/glass.arff

Algorithms

Add new...

Delete selected

J48 -C 0.25 -M 2  
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a  
NaiveBayes

Notes



# Weka Experiment Environment

Setup

Run

Analyse

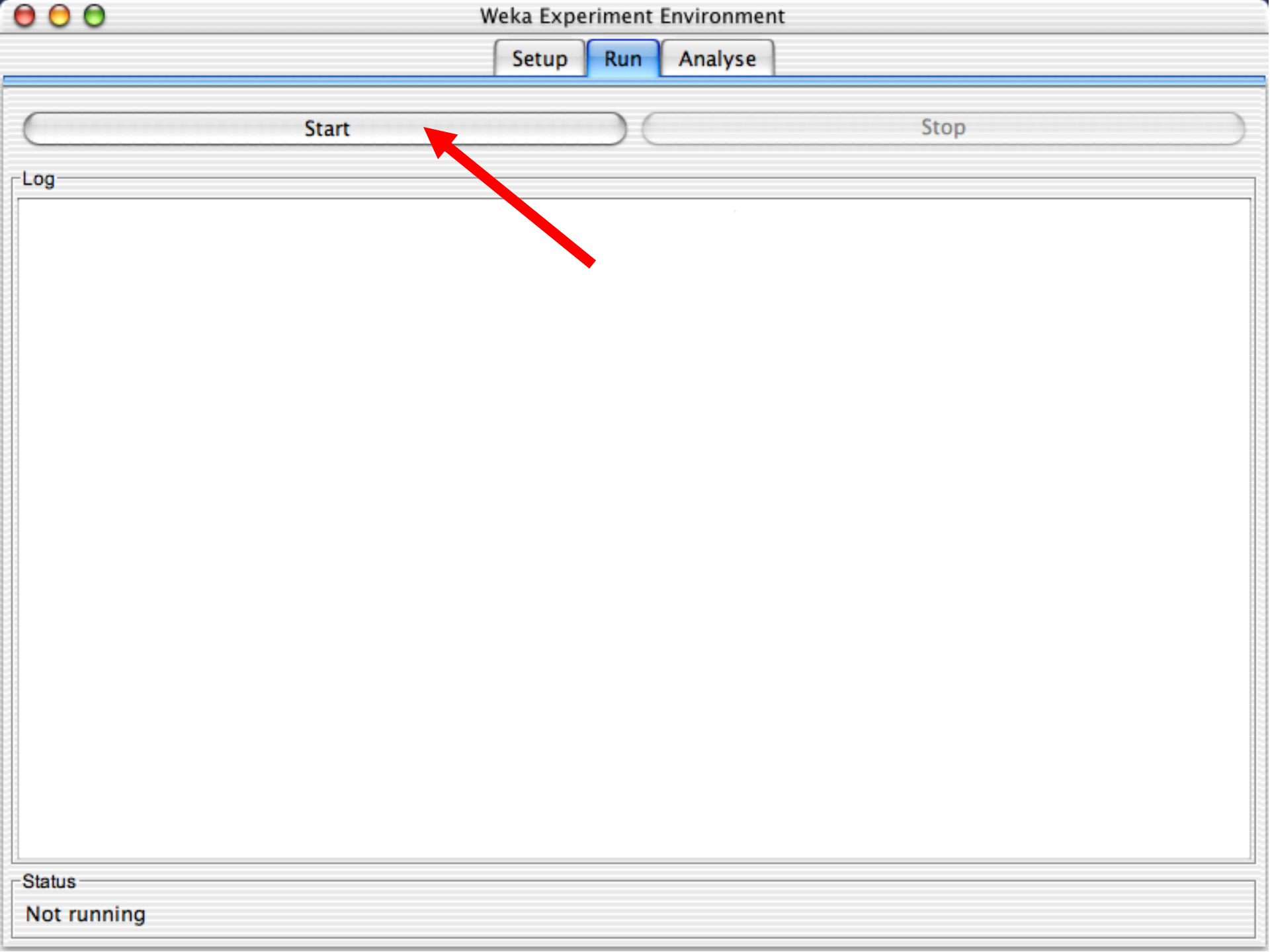
Start

Stop

Log

Status

Not running



# Weka Experiment Environment

Setup

Run

Analyse

Start

Stop

Log

Status

Not running



# Weka Experiment Environment

Setup

Run

Analyse

Start

Stop

Log

10:33:04: Started  
13:41:15: Finished  
13:41:15: There were 0 errors

Status

Not running



# Weka Experiment Environment

Setup

Run

Analyse

Start

Stop

Log

10:33:04: Started  
13:41:15: Finished  
13:41:15: There were 0 errors

Status

Not running





Setup

Run

Analyse

Source

No source

File...

Database...

Experiment

Configure test

Row key fields

Select keys...

Run field



Column key fields

Select keys...

Comparison field



Significance 0.05

Test base

Select base...

Show std. deviations ☐

Perform test

Save output

Result list

Test output



Setup

Run

Analyse

Source

No source

File...

Database...

Experiment

Configure test

Row key fields Run field Column key fields Comparison field Significance Test base Show std. deviations ☐

Result list

Test output



Setup

Run

Analyse

Source

Got 900 results

File...

Database...

Experiment

Configure test

Row key fields Select keys...Run field Key\_RunColumn key fields Select keys...Comparison field Percent\_correctSignificance 0.05Test base Select base...Show std. deviations ☐Perform testSave output

Result list

13:44:17 - Available resultsets

13:44:55 - Percent\_correct - trees.j48.J48 '-C 0

Test output

Analysing: Percent\_correct  
 Datasets: 3  
 Resultsets: 3  
 Confidence: 0.05 (two tailed)  
 Date: 9/9/03 1:44 PM

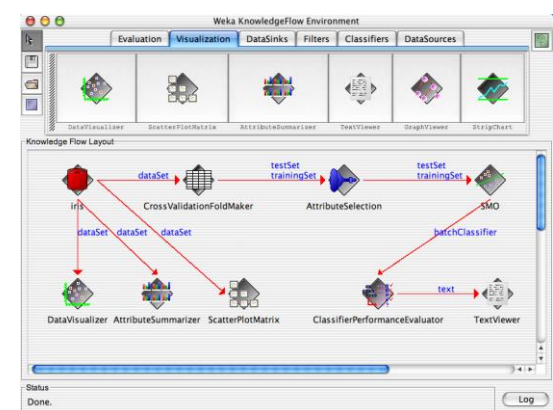
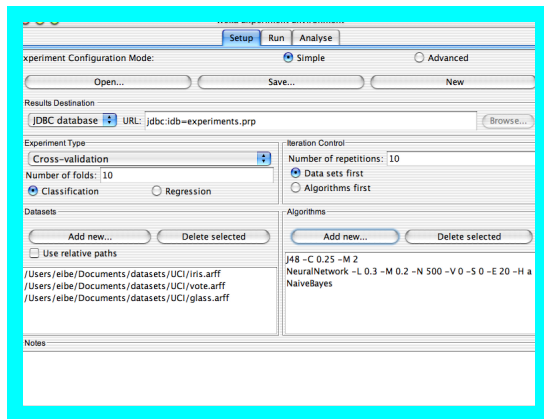
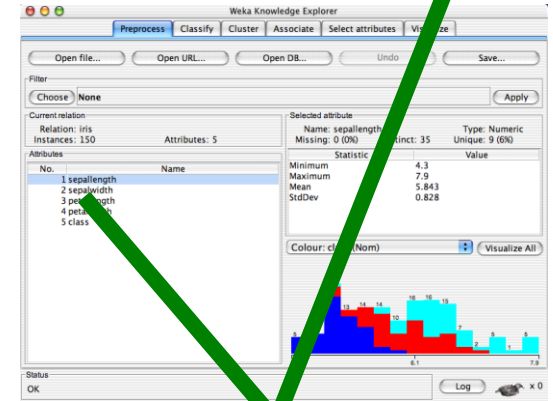
| Dataset | (1) trees.j4 | (2) funct | (3) bayes |
|---------|--------------|-----------|-----------|
| iris    | (100) 94.73  | 96.4      | 95.53     |
| vote    | (100) 96.57  | 94.71 *   | 90.02 *   |
| Glass   | (100) 67.63  | 66.78     | 49.45 *   |

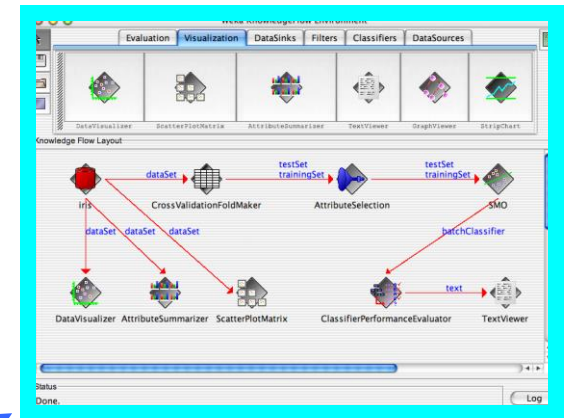
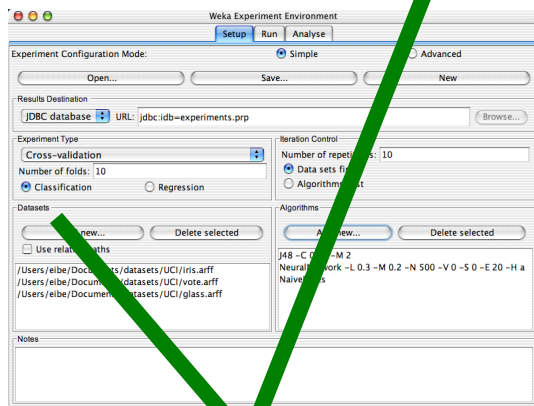
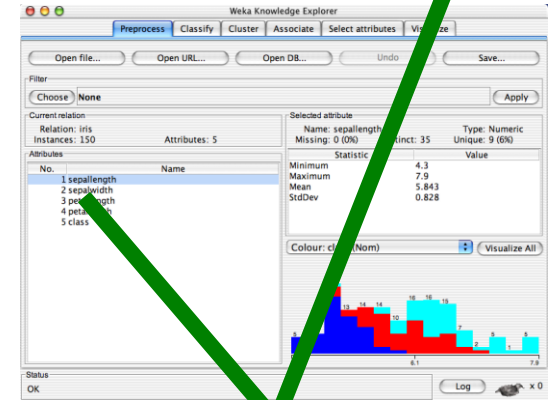
(v/ /\*) | (0/2/1) (0/1/2)

Skipped:

Key:

(1) trees.j48.J48 '-C 0.25 -M 2' -217733168393644444  
 (2) functions.neural.NeuralNetwork '-L 0.3 -M 0.2 -N 500 -V 0 -S 0  
 (3) bayes.NaiveBayes '' 2029074699749330519





# The Knowledge Flow GUI

- New graphical user interface for WEKA
- Java-Beans-based interface for setting up and running machine learning experiments
- Data sources, classifiers, etc. are beans and can be connected graphically
- Data “flows” through components: e.g.,  
“data source” -> “filter” -> “classifier” -> “evaluator”
- Layouts can be saved and loaded again later



# Weka KnowledgeFlow Environment

Evaluation

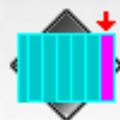
Visualization

DataSinks

Filters

Classifiers

DataSources



TrainTestSplitMaker

ClassAssigner

ClassifierPerformanceEvaluator

IncrementalClassifierEvaluator

PredictionAppender

Knowledge Flow Layout

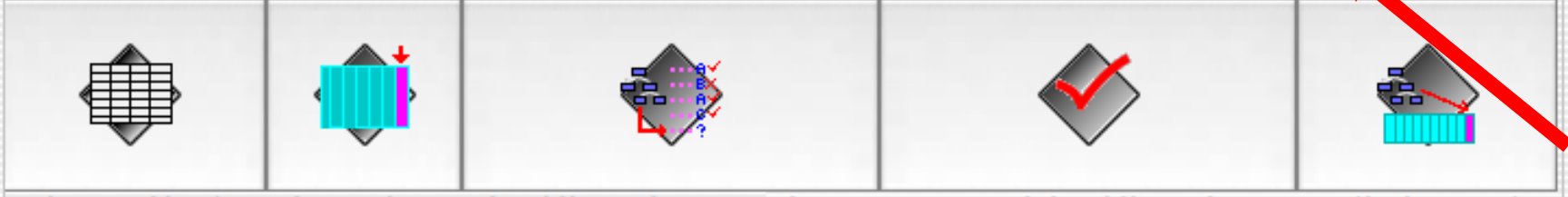
Status

Done.

Log

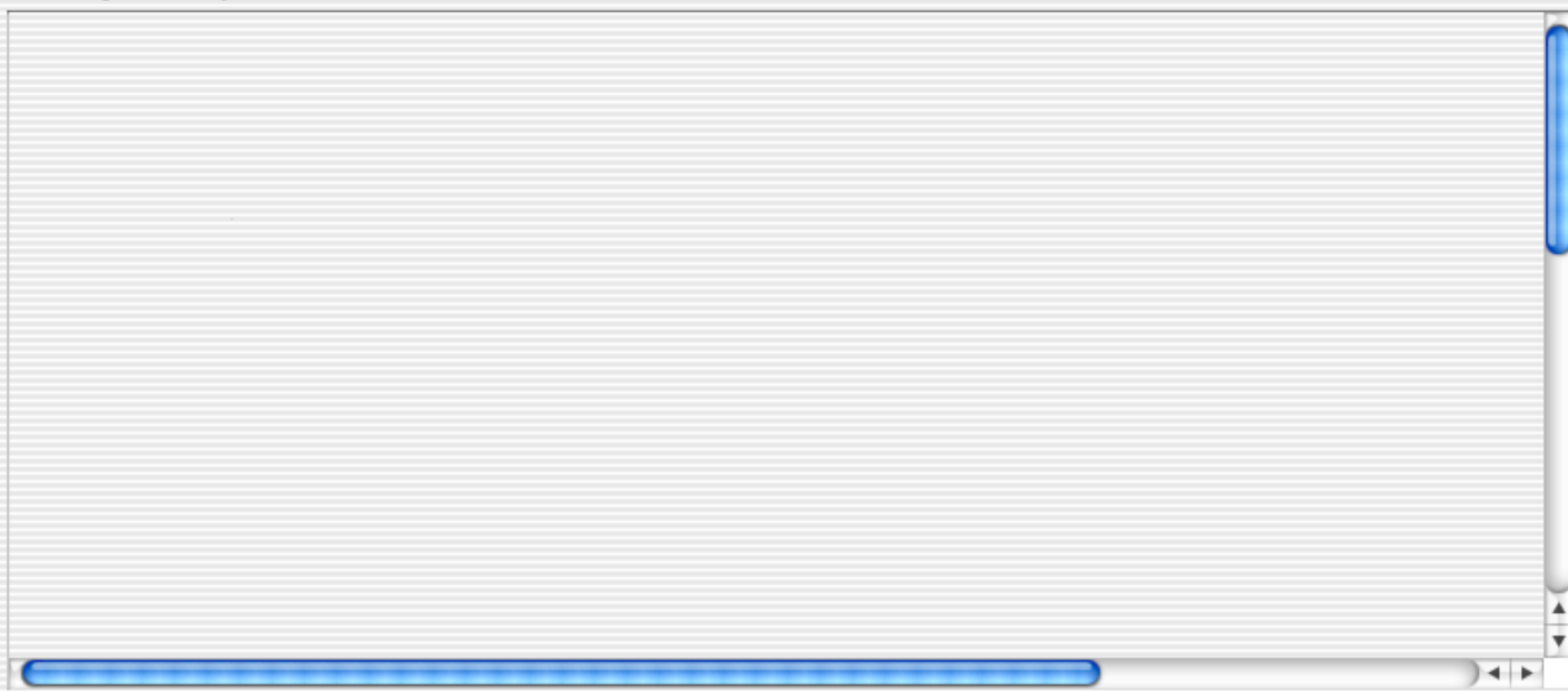
Weka KnowledgeFlow Environment

Evaluation Visualization DataSinks Filters Classifiers DataSources



TrainTestSplitMaker ClassAssigner ClassifierPerformanceEvaluator IncrementalClassifierEvaluator PredictionAppender

Knowledge Flow Layout



Status

Done.

Log



# Weka KnowledgeFlow Environment

Evaluation

Visualization

DataSinks

Filters

Classifiers

**DataSources**



ArffLoader



CSVLoader



C45Loader



SerializedInstancesLoader

## Knowledge Flow Layout



ArffLoader

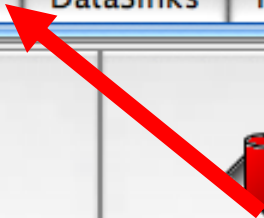
Status

Done.

Log

Weka KnowledgeFlow Environment

Evaluation Visualization DataSinks Filters Classifiers DataSources



ArffLoader CSVLoader C45Loader SerializedInstancesLoader

Knowledge Flow Layout



ArffLoader

Status  
Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

Knowledge Flow Layout



ArffLoader



DataVisualizer

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

Knowledge Flow Layout



ArffLoader



DataVisualizer

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

## Knowledge Flow Layout



Edit

Delete

Configure...

Connections

dataSet

instance

Actions

Start loading



DataVisualizer

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer

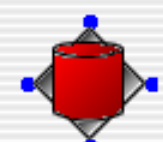


GraphViewer



StripChart

Knowledge Flow Layout



ArffLoader



DataVisualizer

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

Knowledge Flow Layout



ArffLoader

dataSet



DataVisualizer

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



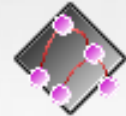
ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

Knowledge Flow Layout



ArffLoader

dataSet



DataVisualizer

Can continue this...

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



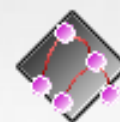
ScatterPlotMatrix



AttributeSummarizer



TextViewer

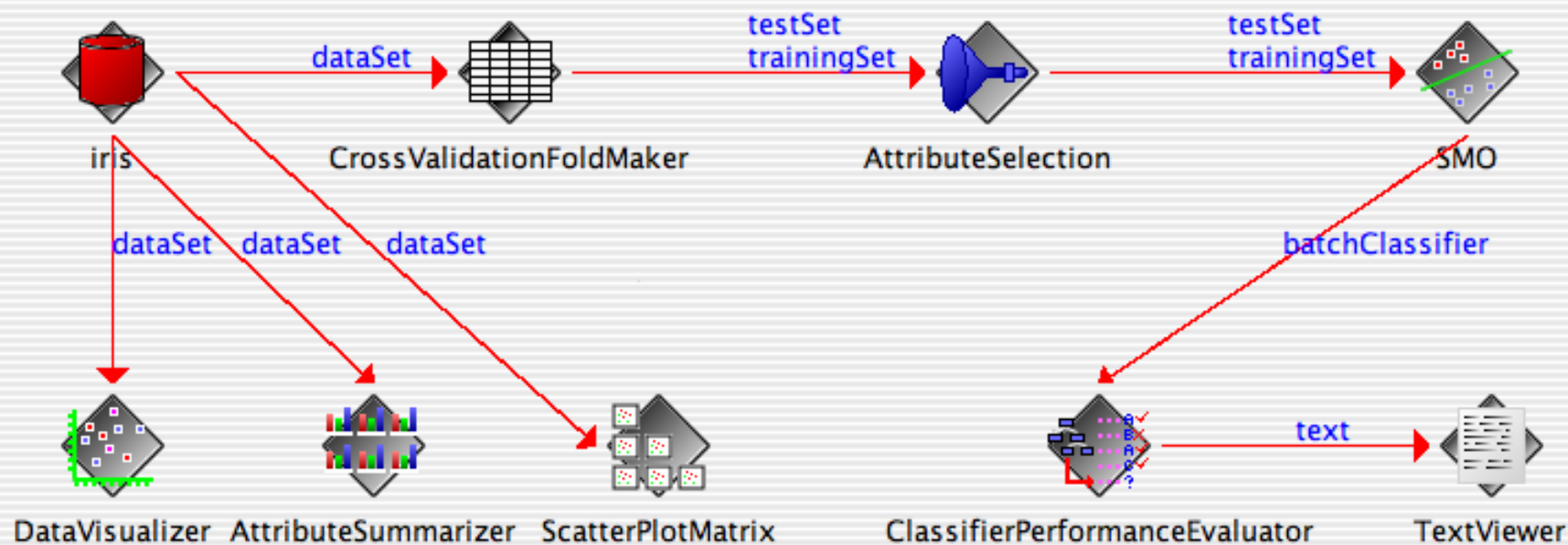


GraphViewer



StripChart

## Knowledge Flow Layout



Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



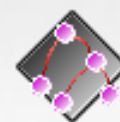
ScatterPlotMatrix



AttributeSummarizer



TextViewer

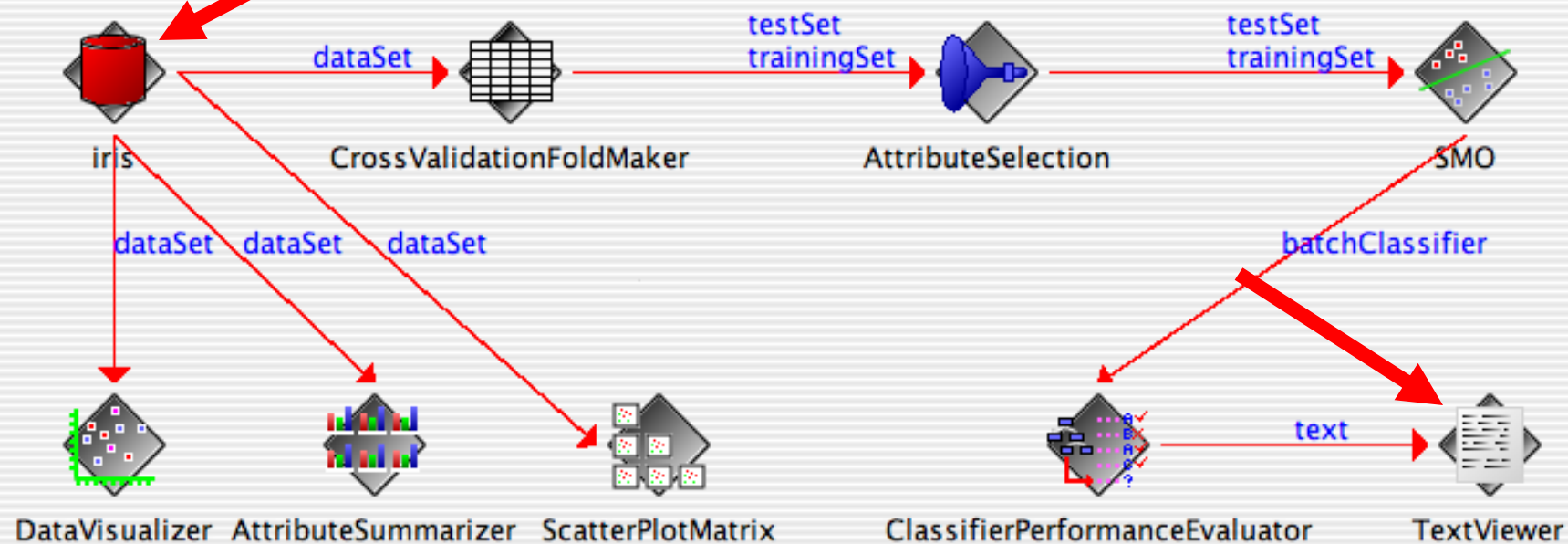


GraphViewer



StripChart

## Knowledge Flow Layout



Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



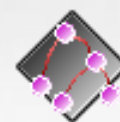
ScatterPlotMatrix



AttributeSummarizer



TextViewer

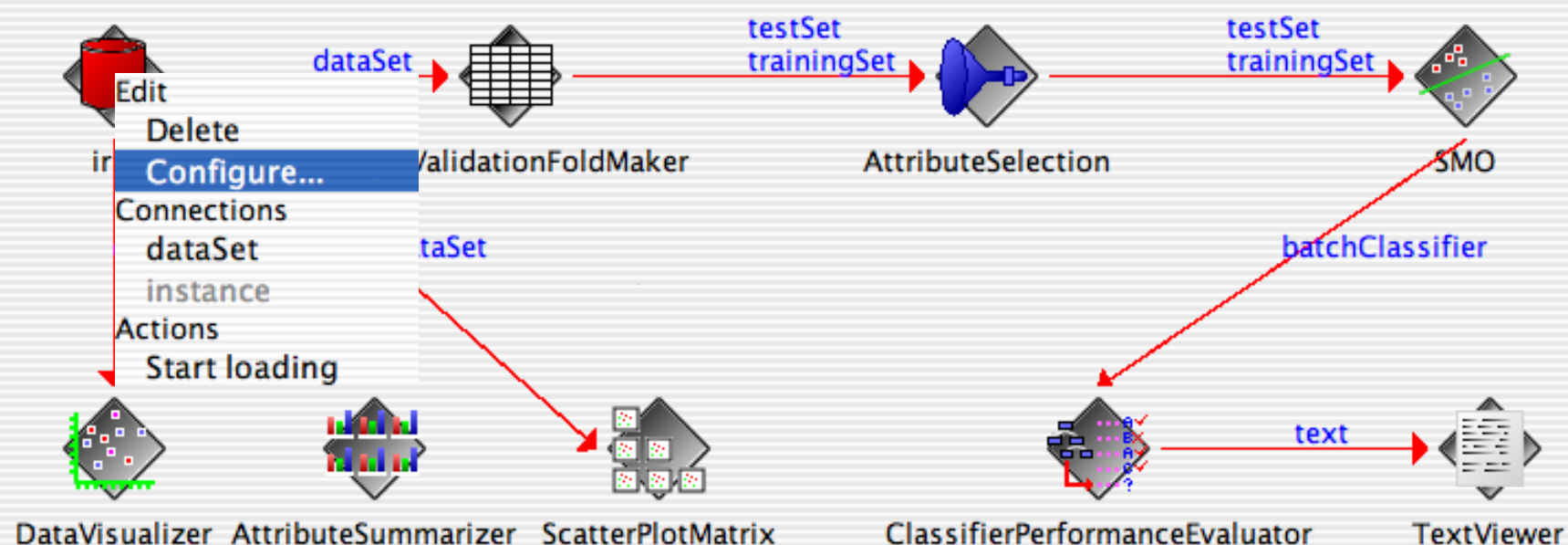


GraphViewer



StripChart

## Knowledge Flow Layout



Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

## Knowledge Flow Layout



dataSet



ir

Choose

weka.core.converters.ArffLoader

file iris.arff

Open...

Save...

OK

Cancel



DataVisualizer



AttributeSummarizer



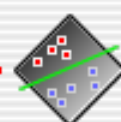
ScatterPlotMatrix

testSet  
trainingSet



AttributeSelection

testSet  
trainingSet



SMO

batchClassifier



ClassifierPerformanceEvaluator

text



TextViewer

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer

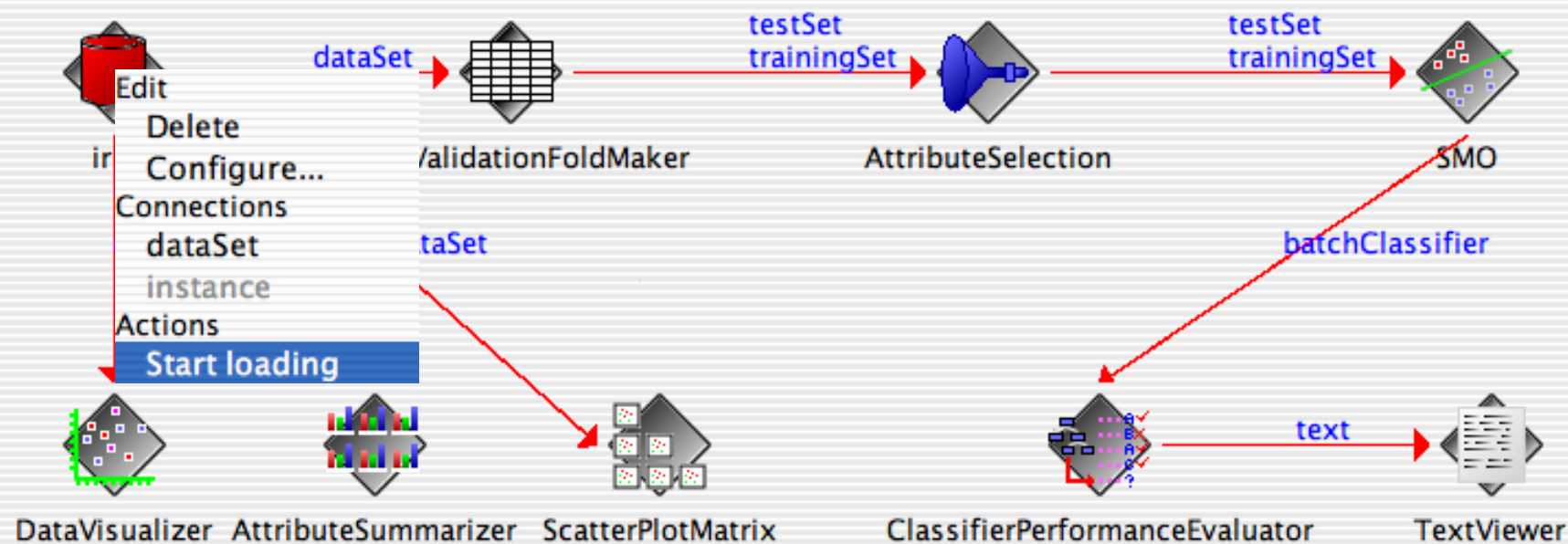


GraphViewer



StripChart

## Knowledge Flow Layout



Status

Done.

Log

Weka KnowledgeFlow Environment

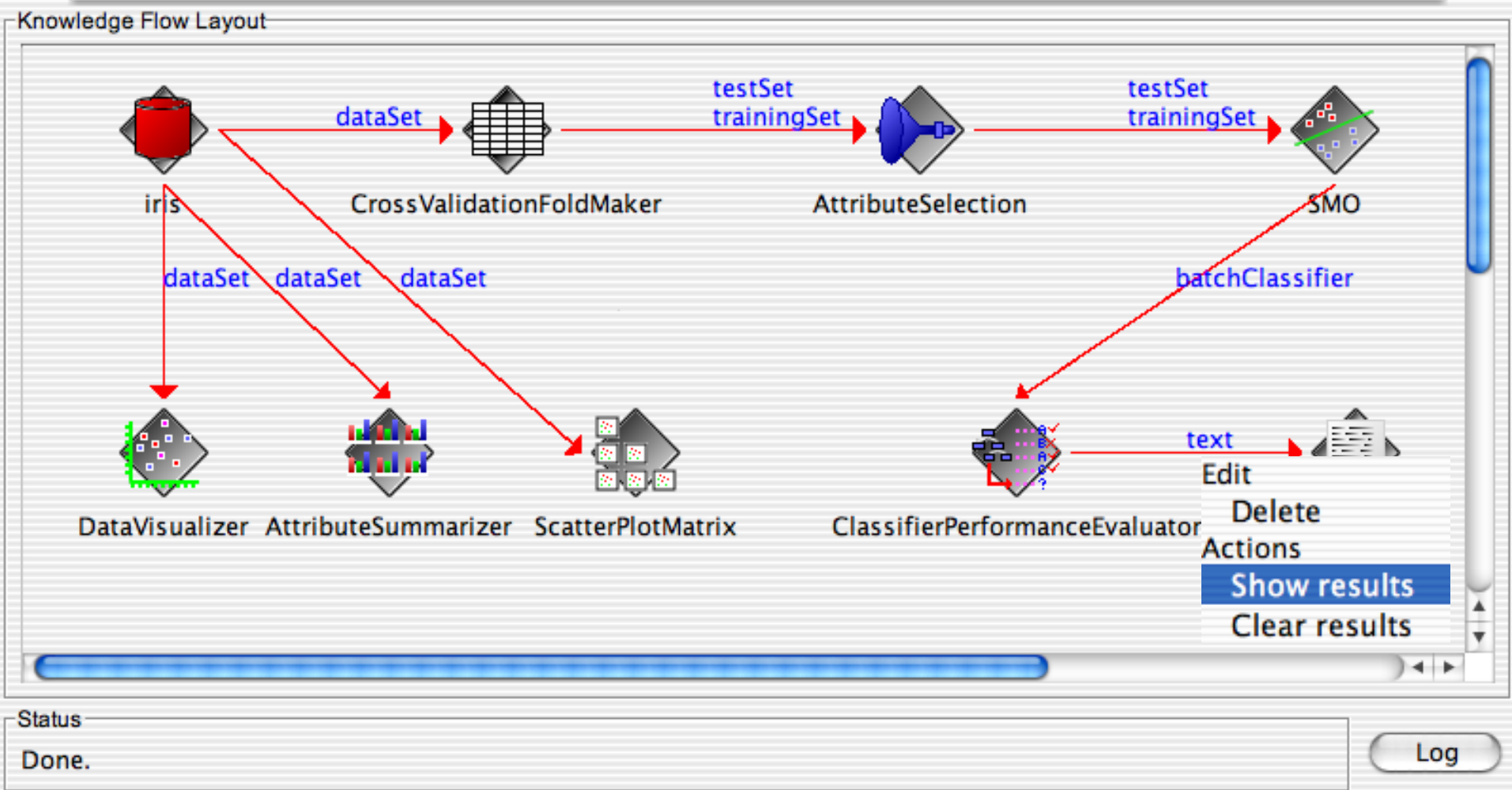
[Evaluation](#)
[Visualization](#)
[DataSinks](#)
[Filters](#)
[Classifiers](#)
[DataSources](#)








DataVisualizer    ScatterPlotMatrix    AttributeSummarizer    TextViewer    GraphViewer    StripChart



Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



DataVisualizer



ScatterPlotMatrix



AttributeSummarizer



TextViewer



GraphViewer



StripChart

Knowledge Flow Layout

## Text Viewer

Result list

09:59:02 - SMO

Text

|                                  |        |    |   |
|----------------------------------|--------|----|---|
| Correctly Classified Instances   | 144    | 96 | % |
| Incorrectly Classified Instances | 6      | 4  | % |
| Kappa statistic                  | 0.94   |    |   |
| Mean absolute error              | 0.2311 |    |   |
| Root mean squared error          | 0.288  |    |   |
| Relative absolute error          | 52     | %  |   |
| Root relative squared error      | 58.704 | %  |   |
| Total Number of Instances        | 150    |    |   |

Data

Status

Done.

Log

Evaluation

Visualization

DataSinks

Filters

Classifiers

DataSources



Logistic



VotedPerceptron



Winnow

lazy



IB1



IBk



KStar



LBR



LWL

meta



AdaBoostM1



AdditiveRec



Knowledge Flow Layout



ArffLoader

instance



IBk

incrementalClassifier

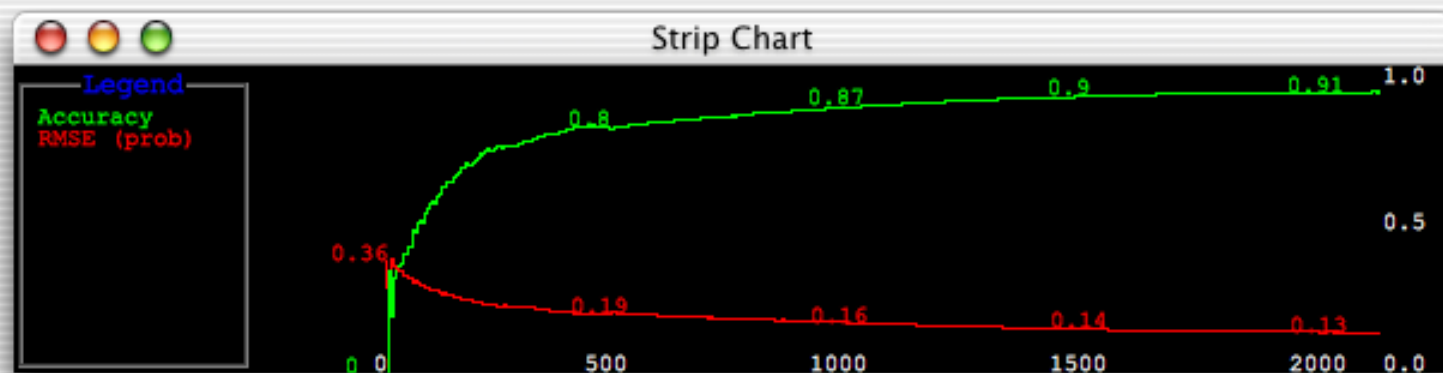


IncrementalClassifierEvaluator

chart



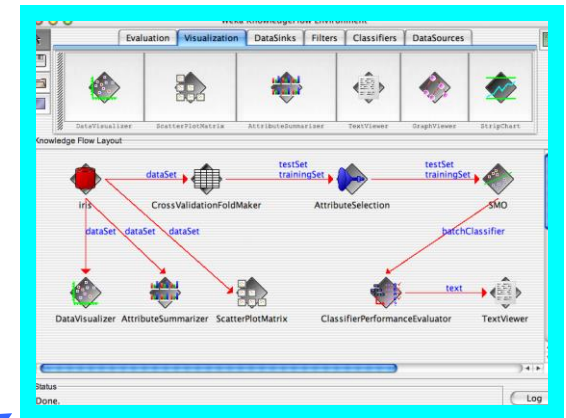
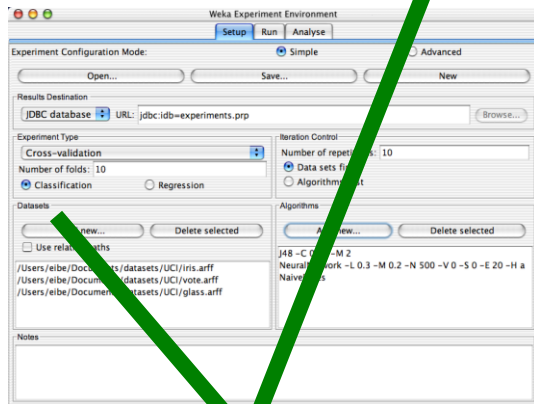
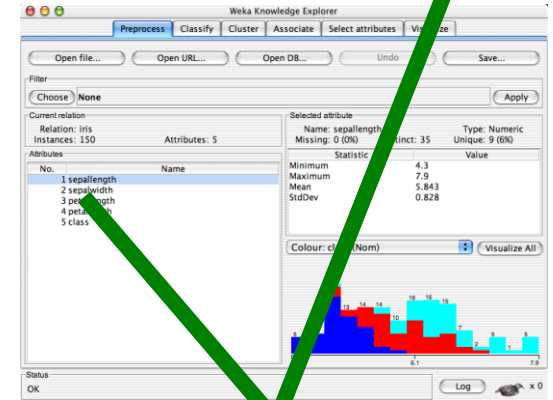
StripChart

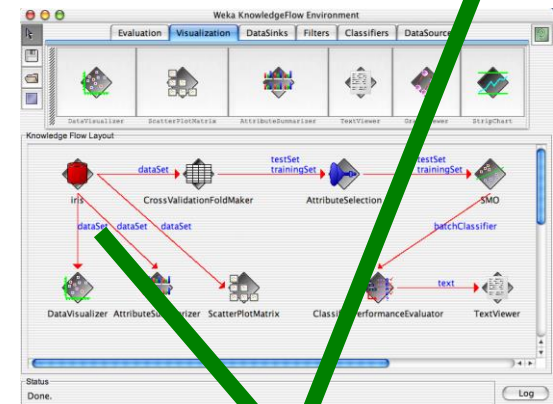
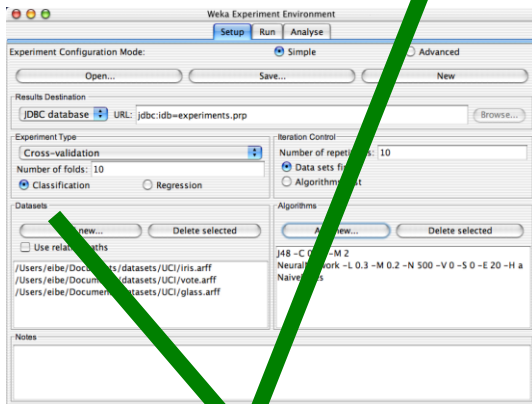
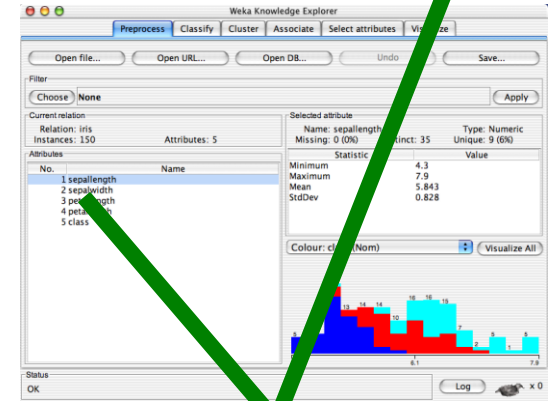


Status

Done.

Log





# Conclusion: try it yourself!

- WEKA is available at  
<http://www.cs.waikato.ac.nz/ml/weka>
- Also has a list of projects based on WEKA
- WEKA contributors:

Abdelaziz Mahoui, Alexander K. Seewald, Ashraf M. Kibriya, Bernhard Pfahringer , Brent Martin, Peter Flach, Eibe Frank ,Gabi Schmidberger ,Ian H. Witten , J. Lindgren, Janice Boughton, Jason Wells, Len Trigg, Lucio de Souza Coelho, Malcolm Ware, Mark Hall ,Remco Bouckaert , Richard Kirkby, Shane Butler, Shane Legg, Stuart Inglis, Sylvain Roy, Tony Voyle, Xin Xu, Yong Wang, Zhihai Wang