

# An Introduction to WEKA Explorer

In part from: Yizhou Sun  
2008

# What is WEKA?

- **Waikato Environment for Knowledge Analysis**
  - A data mining/machine learning tool developed by Department of Computer Science, University of Waikato, New Zealand.
  - Weka is also a bird found only on the islands of New Zealand.



# How does it works?

- First, you select a dataset and a Machine learning algorithm
- You can manipulate the dataset in several ways, as we will see.
- When dataset is ready, you select a ML algorithm from the list, and adjust learning parameters, as we will see
- When you run a ML algorithm, the system will:
  1. Split the data set into training and testing subsets;
  2. Learn a classification function  $C(x)$  based on examples in the training set;
  3. Classify instances  $x$  in the test set based on the learned function  $C(x)$ ;
  4. Measure the performances by comparing the generated classifications with the “ground truth” in the test set.

# Download and Install WEKA

- Website:

<http://www.cs.waikato.ac.nz/~ml/weka/index.html>

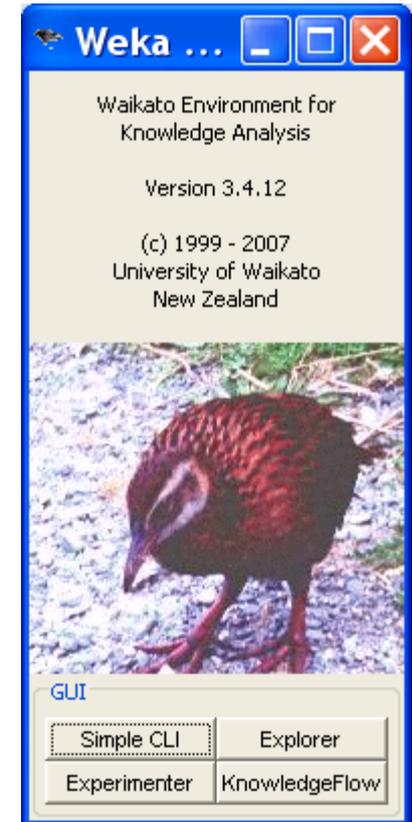
- Support multiple platforms (written in java):
  - Windows, Mac OS X and Linux

# Main Features

- 49 data preprocessing tools
- 76 classification/regression algorithms
- 8 clustering algorithms
- 3 algorithms for finding association rules
- 15 attribute/subset evaluators + 10 search algorithms for feature selection

# Main GUI

- Three graphical user interfaces
  - “The Explorer” (exploratory data analysis)
  - “The Experimenter” (experimental environment)
  - “The KnowledgeFlow” (new process model inspired interface)
- Simple CLI- provides users without a graphic interface option the ability to execute commands from a terminal window



# Explorer

- **The Explorer:**
  - Preprocess data
  - Classification
  - Clustering
  - Association Rules
  - Attribute Selection
  - Data Visualization
- References and Resources

# 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 **Java DataBase Connectivity**)
- Pre-processing tools in WEKA are called “filters”
- WEKA contains filters for:
  - Discretization, normalization, resampling, attribute selection, transforming and combining attributes, ...

# 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

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

Empty list area for attributes.

Empty list area for selected attributes.



Visualize All

Status

Welcome to the Weka Knowledge Explorer

Log



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

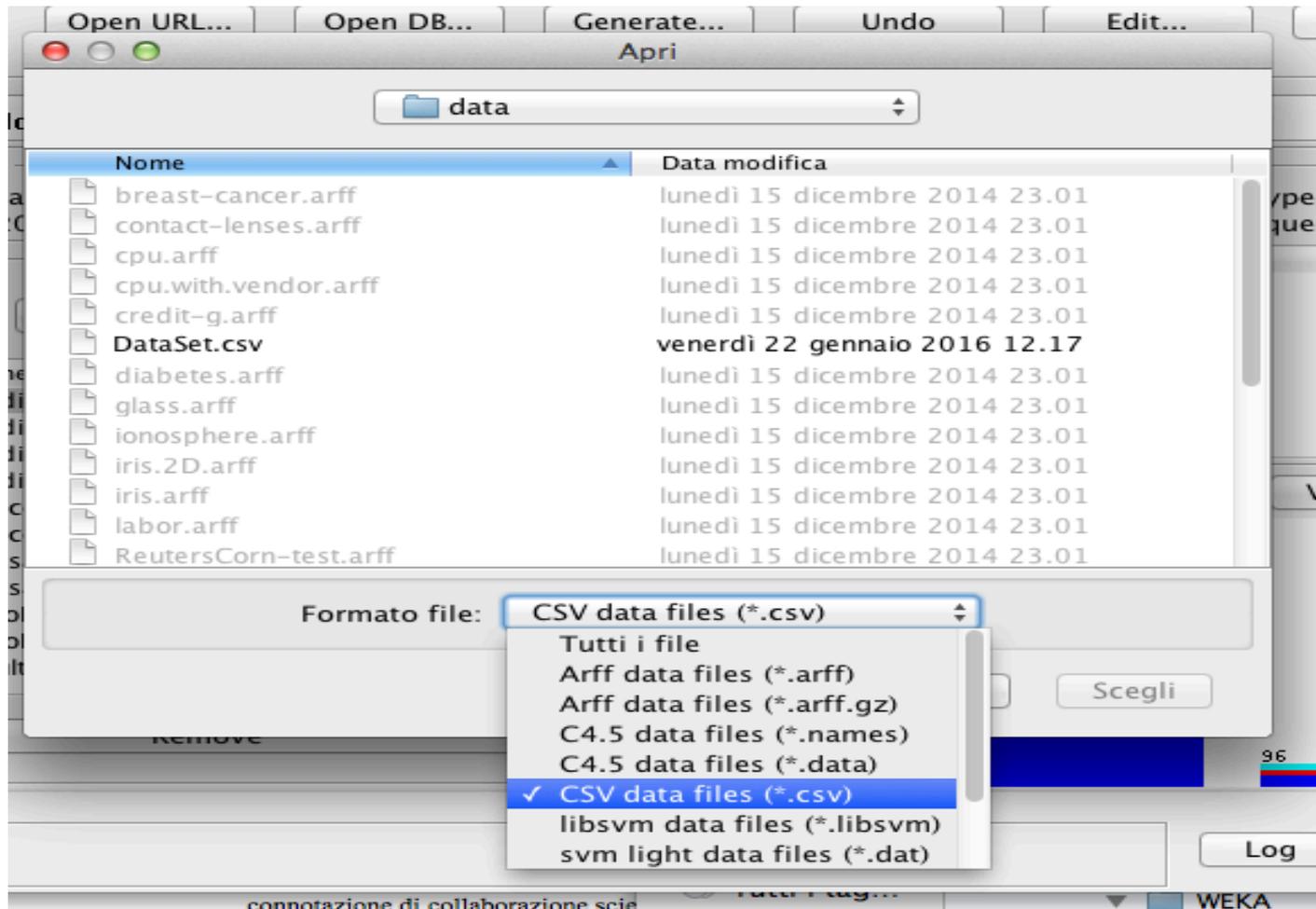
Status

Welcome to the Weka Knowledge Explorer

Log



# You can either open arff file or convert from other formats



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

Current relation

Relation: None

Instances: None

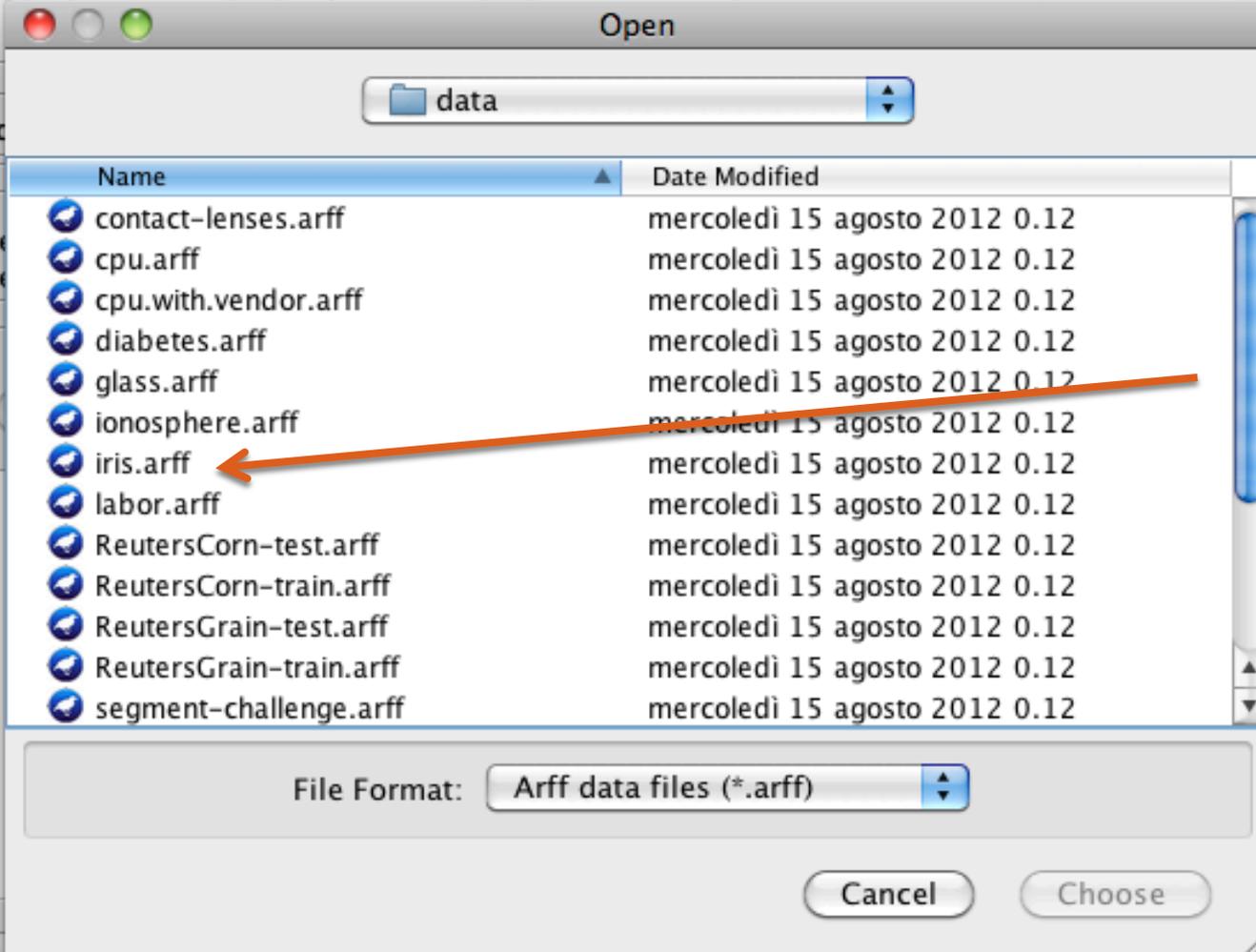
Attributes

All

Apply

Type: None  
Unique: None

Visualize All



File Format:

Arff data files (\*.arff)

Cancel

Choose

Status

Welcome to the Weka Explorer

Log

 x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

No

Current relation

Relation: iris

Instances: 150

Attributes

All

No.	Name
1	<input type="checkbox"/> sepal
2	<input type="checkbox"/> sepal
3	<input type="checkbox"/> petal
4	<input type="checkbox"/> petal
5	<input type="checkbox"/> class

Open

data

Name	Date Modified
<input checked="" type="checkbox"/> contact-lenses.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> cpu.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> cpu.with.vendor.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> diabetes.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> glass.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> ionosphere.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> iris.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> labor.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> ReutersCorn-test.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> ReutersCorn-train.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> ReutersGrain-test.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> ReutersGrain-train.arff	mercoledì 15 agosto 2012 0.12
<input checked="" type="checkbox"/> segment-challenge.arff	mercoledì 15 agosto 2012 0.12

File Format: Arff data files (\*.arff)

Cancel

Choose

Visualize All

10

7

4.3

6.1

7.9

Status

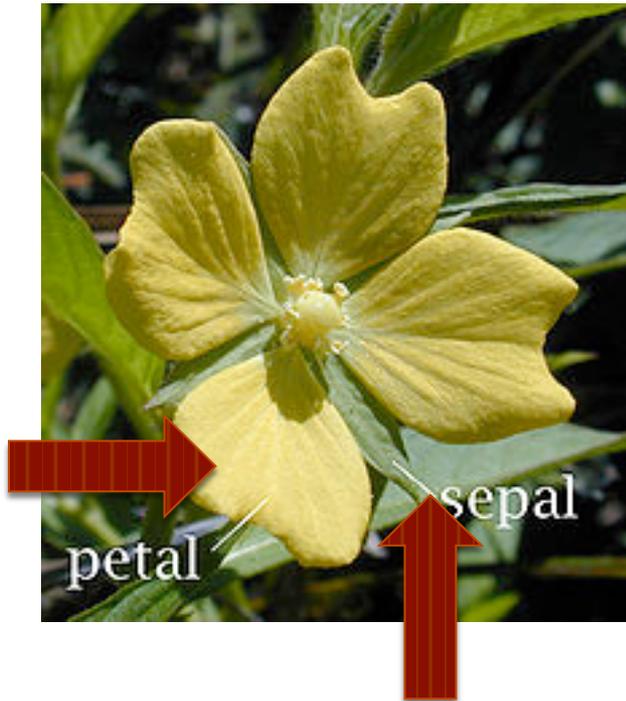
OK

Log



# IRIS dataset

- 150 instances of IRIS (a flower)
- 5 attributes, one is the classification  $c(x)$
- 3 classes: iris setosa, iris versicolor, iris virginica



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

None

For any selected attribute you can get statistics

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Selected attribute

Name: sepallength

Type: Numeric

Missing: 0 (0%)

Distinct: 35

Unique: 9 (6%)

Attributes

All

None

Invert

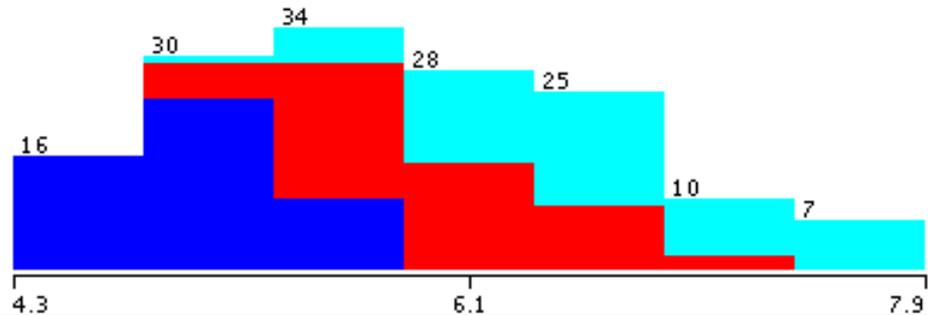
Pattern

- | No. | Name  |
|-----|---|
| 1   | <input checked="" type="checkbox"/> sepallength |
| 2   | <input type="checkbox"/> sepalwidth             |
| 3   | <input type="checkbox"/> petallength            |
| 4   | <input type="checkbox"/> petalwidth             |
| 5   | <input type="checkbox"/> class                  |

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Class: class (Nom)

Visualize All



Remove

Status

OK

Log



# Attribute data

- Min, max and average value of attributes
- distribution of values :number of items for which:

$$a_i = v_j \mid a_i \in A, v_j \in V$$

- class: distribution of attribute values in the classes
- The class (e.g.  $C(x)$ , the classification function to be learned) is by default THE LAST ATTRIBUTE of the list.

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

None

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

All

None

Invert

Pattern

No.	Name
1	<input checked="" type="checkbox"/> sepallength
2	<input type="checkbox"/> sepalwidth
3	<input type="checkbox"/> petallength
4	<input type="checkbox"/> petalwidth
5	<input type="checkbox"/> class

Remove

Selected attribute

Name: sepallength

Type: Numeric

Missing: 0 (0%)

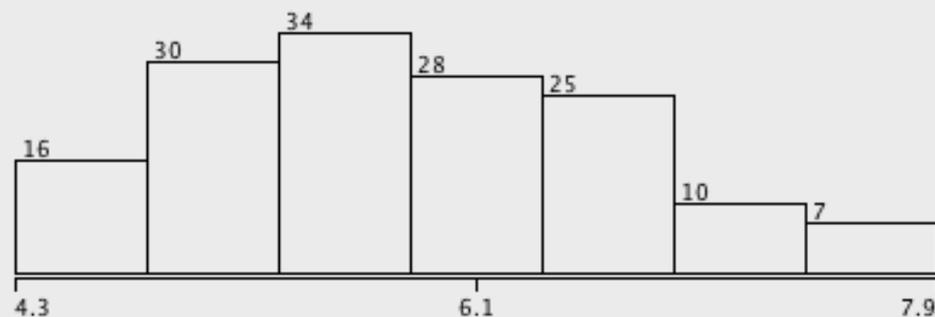
Distinct: 35

Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Class: sepallength (Num)

Visualize All



Status

OK

Log



x 0

Filter: Choose **None** Apply

Current relation  
 Relation: iris  
 Instances: 150  
 Attributes: 5

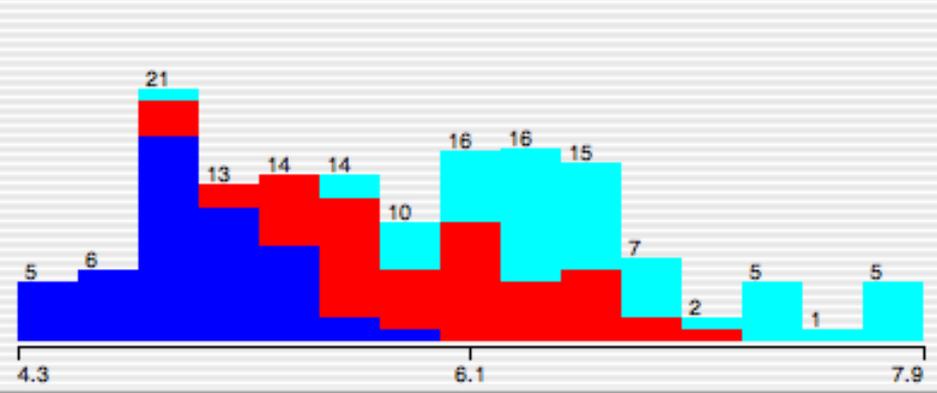
Selected attribute  
 Name: sepallength  
 Missing: 0 (0%)  
 Distinct: 35  
 Type: Numeric  
 Unique: 9 (6%)

Attributes

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

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Colour: class (Nom) Visualize All



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

Selected attribute

Name: class  
Missing: 0 (0%)  
Distinct: 3  
Type: Nominal  
Unique: 0 (0%)

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

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 None

Apply

Current relation

Relation: iris  
Instances: 150

Attributes: 5

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: class  
Missing: 0 (0%)  
Distinct: 3  
Type: Nominal  
Unique: 0 (0%)

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

Colour: class (Nom)

Visualize All

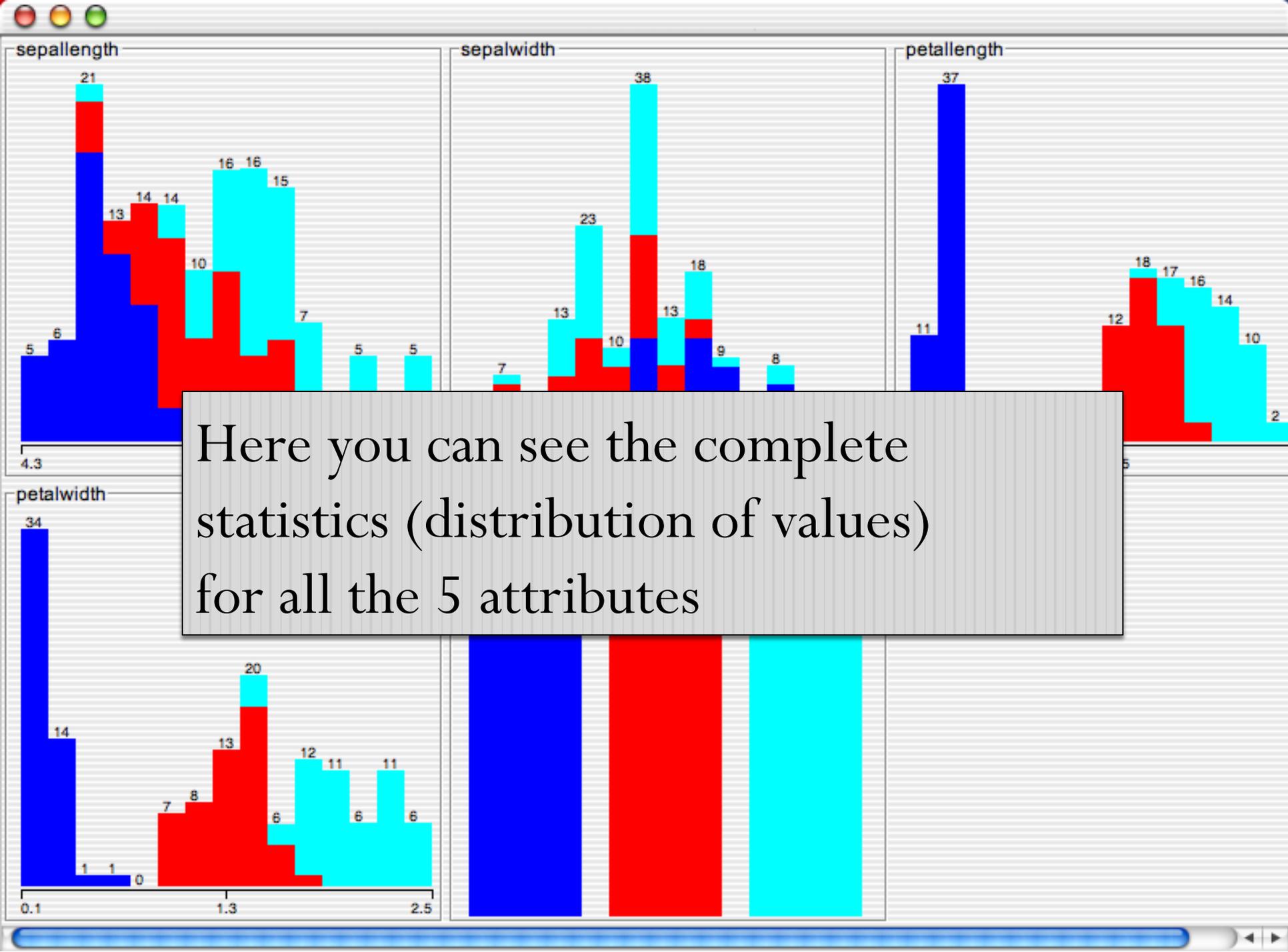


Status

OK

Log

 x 0



# Filtering attributes

- Once the initial data has been selected and loaded the user can select options for **refining the experimental data**.
- The options in the preprocess window include selection of **optional filters** to apply and the user can **select or remove different attributes** of the data set as necessary to identify specific information (or even write a regex in Perl).
- The user can modify the attribute selection and change the relationship among the different attributes by deselecting different choices from the original data set.
- There are many different filtering options available within the preprocessing window and the user can select the different options based on need and type of data present.

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	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

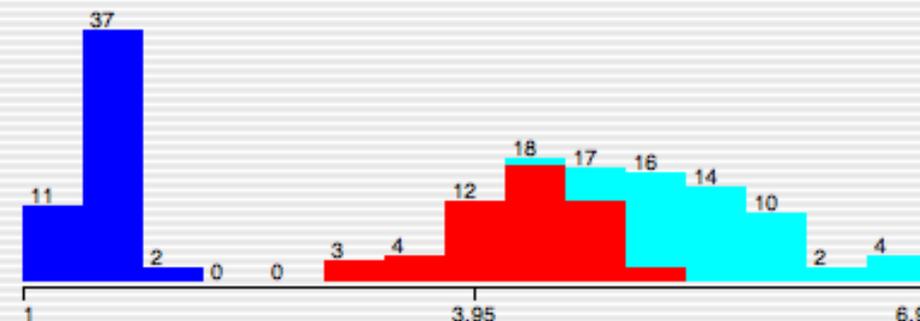
Selected attribute

Name: petallength  
Missing: 0 (0%)      Distinct: 43      Type: Numeric  
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 None

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

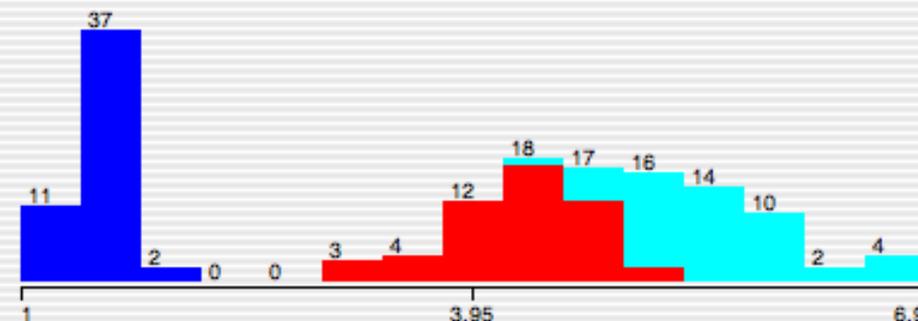
Distinct: 43

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

- weka
  - filters
    - unsupervised
      - attribute
      - instance

Apply

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

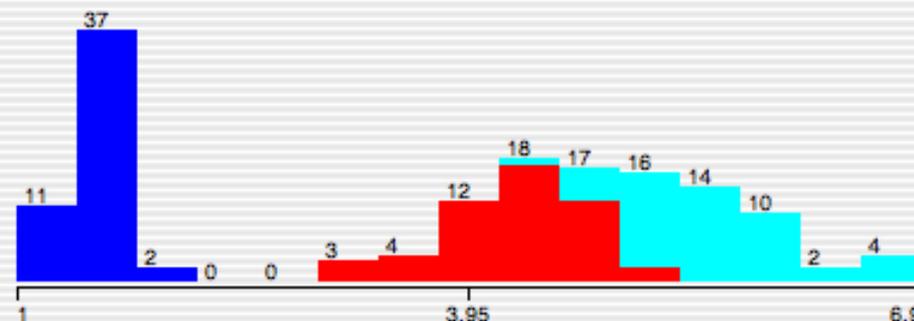
Distinct: 43

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...

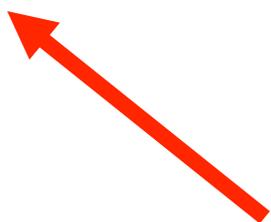
Open DB...

Undo

Save...

Filter

- weka
  - filters
    - unsupervised
      - attribute
      - instance



Apply

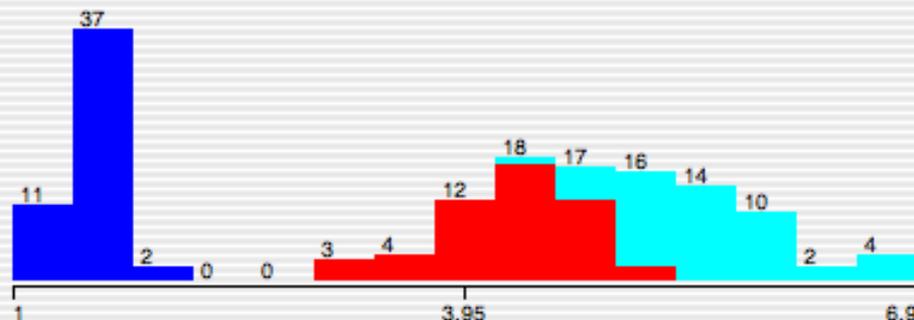
Selected attribute

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



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Undo

Save...

Filter

- weka
  - filters
    - unsupervised
      - attribute
        - Add
        - AddCluster
        - AddExpression
        - AddNoise
        - Copy
        - Discretize
        - FirstOrder
        - MakeIndicator
        - MergeTwoValues
        - NominalToBinary
        - Normalize
        - NumericToBinary
        - NumericTransform
        - Obfuscate
        - PKIDiscretize
        - Remove
        - RemoveType

Apply

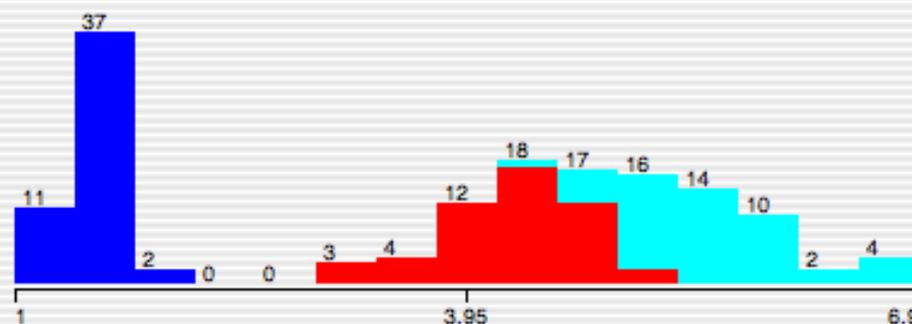
Selected attribute

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

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

Distinct: 43

Unique: 10 (7%)

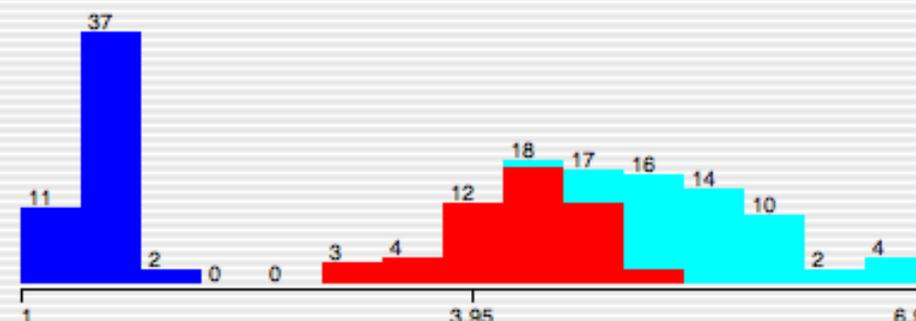
Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

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	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

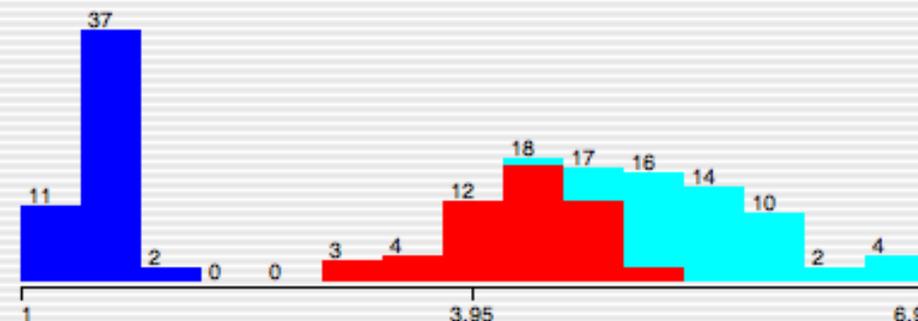
Distinct: 43

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**

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

: Numeric  
: 10 (7%)

e

attributeIndices first-last

bins 10

findNumBins False

invertSelection False

makeBinary False

useEqualFrequency False

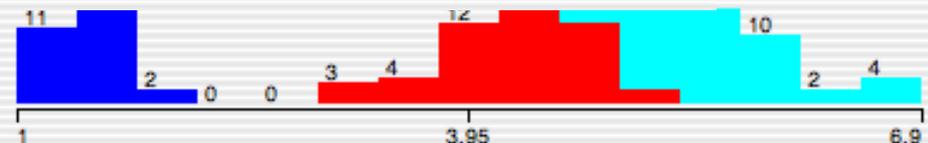
Visualize All

Open...

Save...

OK

Cancel



Status

OK

Log



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: !

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
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

: Numeric  
: 10 (7%)

attributeIndices first-last

bins 10

findNumBins False

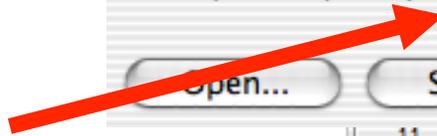
invertSelection False

False

useEqualFrequency False

Visualize All

Discretizes in 10 bins of equal frequency

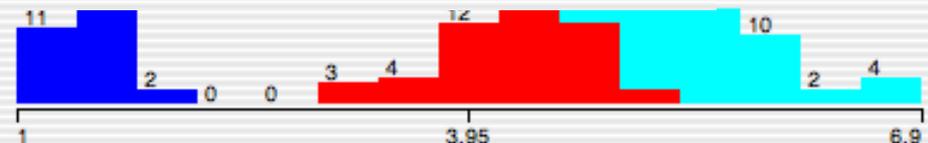


Open...

Save...

OK

Cancel



Status

OK

Log



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	sepalength
2	sepalwidth
3	petallength
4	petalwidth
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

: Numeric  
: 10 (7%)

e

attributeIndices first-last

bins 10

findNumBins False

invertSelection False

False

useEqualFrequency True

Visualize All

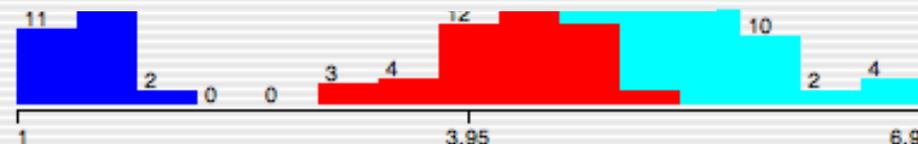
Discretizes in 10 bins of equal frequency

Open...

Save...

OK

Cancel



Status

OK

Log



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: !

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
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

: Numeric  
: 10 (7%)

attributeIndices first-last

bins 10

findNumBins False

invertSelection False

False

useEqualFrequency True

Visualize All

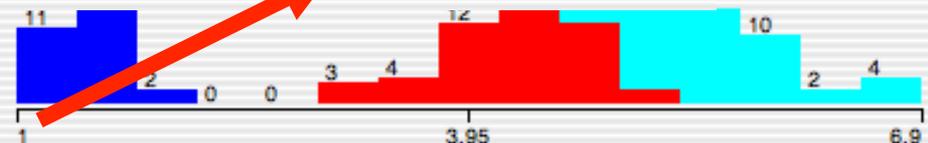
Discretizes in 10 bins of equal frequency

Open...

Save...

OK

Cancel



Status

OK

Log



x 0

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

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

Distinct: 43

Unique: 10 (7%)

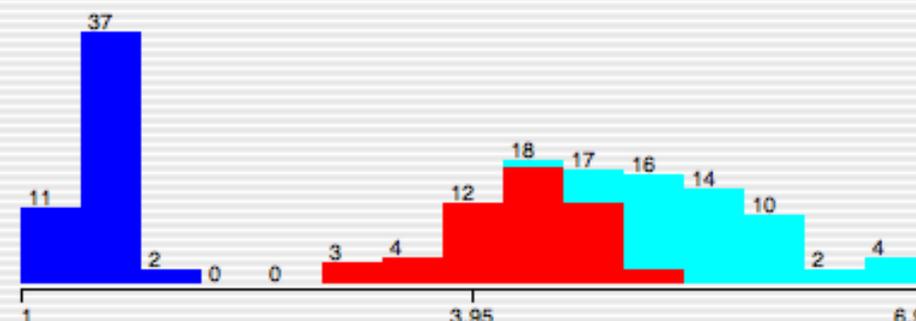
Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

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 -F -B 10 -R first-last

Apply

Current relation

Relation: iris

Instances: 150

Attributes: 5

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: petallength

Type: Numeric

Missing: 0 (0%)

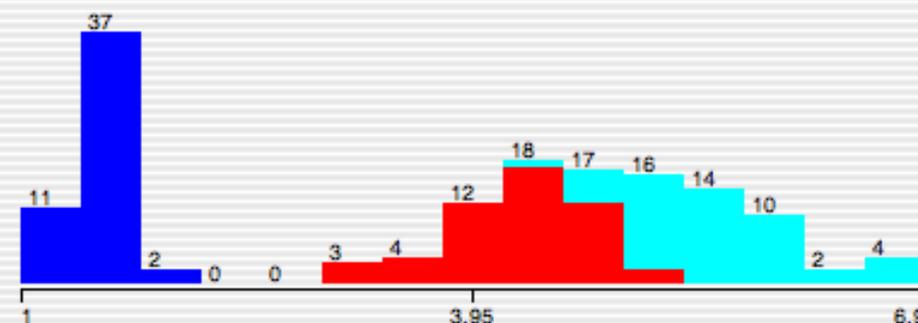
Distinct: 43

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 -F -B 10 -R first-last

Apply

Current relation

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

Instances: 150

Attributes: 5

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: petallength

Type: Nominal

Missing: 0 (0%)

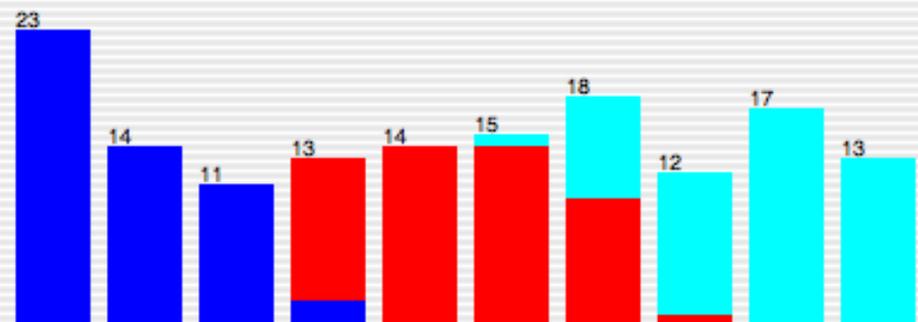
Distinct: 10

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

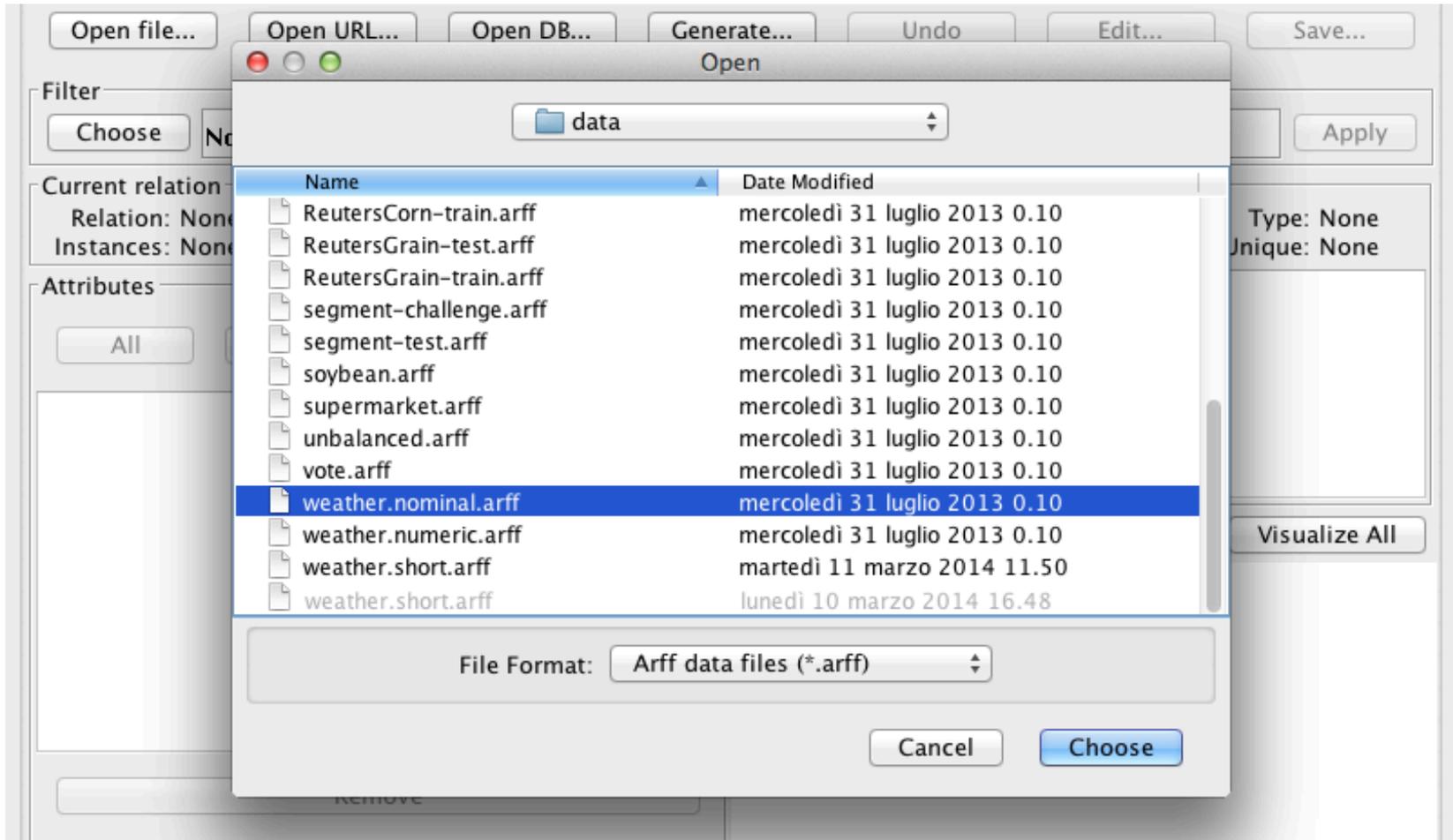
x 0

**WILL SEE MORE ON FILTERING  
DURING FIRST LAB!!**

# Explorer: building “classifiers”

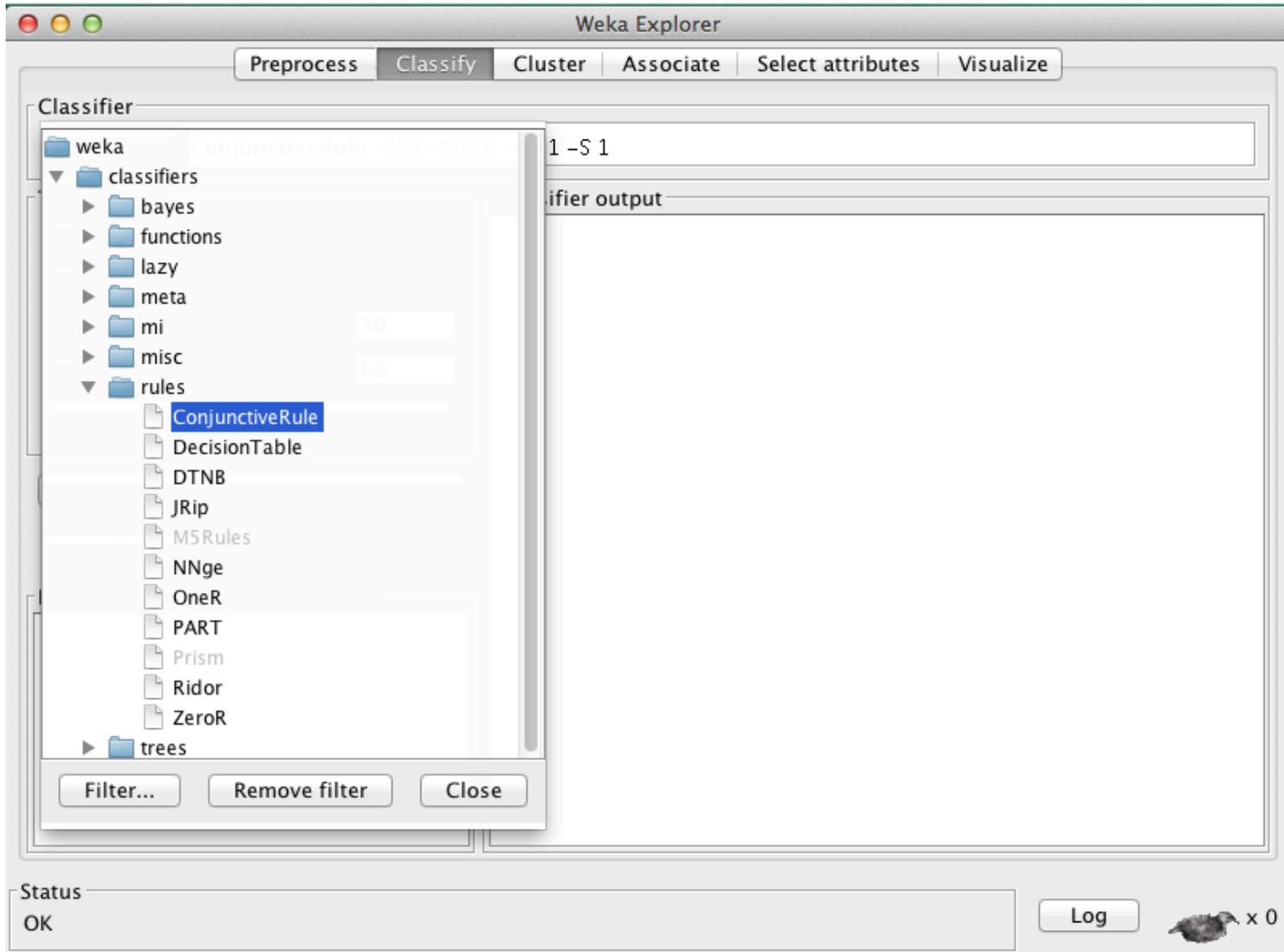
- “Classifiers” in WEKA are machine learning algorithms for predicting nominal or numeric values of a selected attribute (e.g. the CLASS attribute in the IRIS file)
- Implemented learning algorithms include:
  - Conjunctive rules, decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes’ nets, ...
  - Most, but not all, the algorithms that we will present in this course (e.g. no genetic or reinforcement algorithms)

# Explore Conjunctive Rules learner

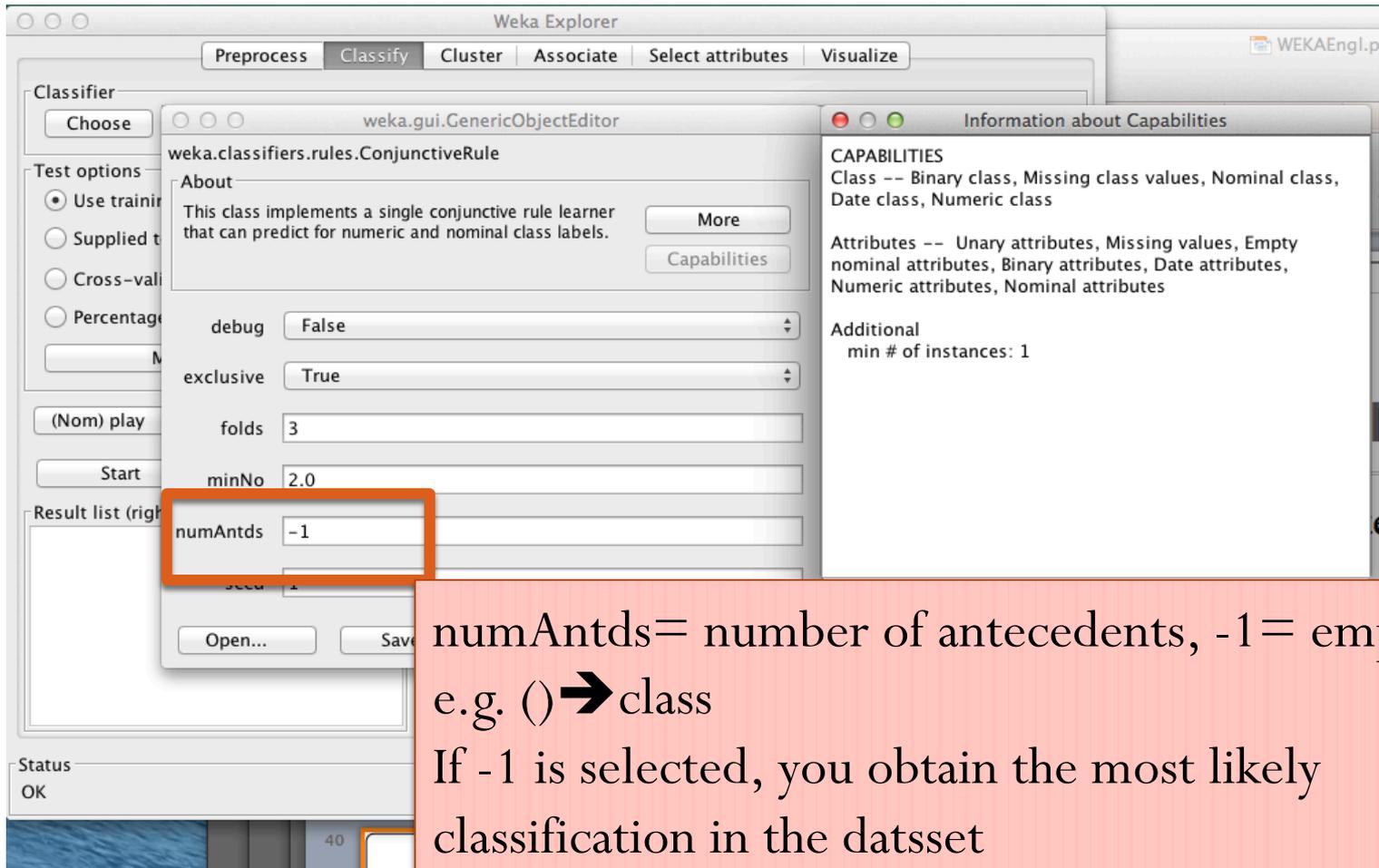


Need a simple dataset with few attributes , let's select the weather dataset

# Select a Classifier



# Right-click to select parameters



The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. A 'weka.gui.GenericObjectEditor' window is open for the 'weka.classifiers.rules.ConjunctiveRule' classifier. The 'About' section states: 'This class implements a single conjunctive rule learner that can predict for numeric and nominal class labels.' The 'Capabilities' section lists: 'CAPABILITIES: Class -- Binary class, Missing class values, Nominal class, Date class, Numeric class' and 'Attributes -- Unary attributes, Missing values, Empty nominal attributes, Binary attributes, Date attributes, Numeric attributes, Nominal attributes'. The 'Additional' section shows 'min # of instances: 1'. The 'numAntds' parameter is highlighted with an orange box and set to -1. A text box explains: 'numAntds = number of antecedents, -1 = empty rule e.g. () → class. If -1 is selected, you obtain the most likely classification in the dataset'.

# Select numAntds=10

The image shows the Weka Explorer interface with the 'Classify' tab selected. A dialog box titled 'weka.gui.GenericObjectEditor' is open, showing the configuration for the 'weka.classifiers.rules.ConjunctiveRule' classifier. The 'numAntds' property is set to 10. Three orange arrows point to the 'Start' button in the background, the 'numAntds' field, and the 'OK' button in the dialog.

Weka Explorer

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

Classifier: Choose

Test options:

- Use training data
- Supplied test data
- Cross-validation
- Percentage split

(Nom) only

Start

Result list (right)

weka.gui.GenericObjectEditor

weka.classifiers.rules.ConjunctiveRule

About

This class implements a single conjunctive rule learner that can predict for numeric and nominal class labels.

More

Capabilities

debug: False

exclusive: False

folds: 3

minNo: 2.0

numAntds: 10

seed: 1

Open... Save... OK Cancel

# Select training method

Weka Explorer

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

Classifier

Choose **ConjunctiveRule -N 3 -M 2.0 -P -1 -S 1 -E**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) play

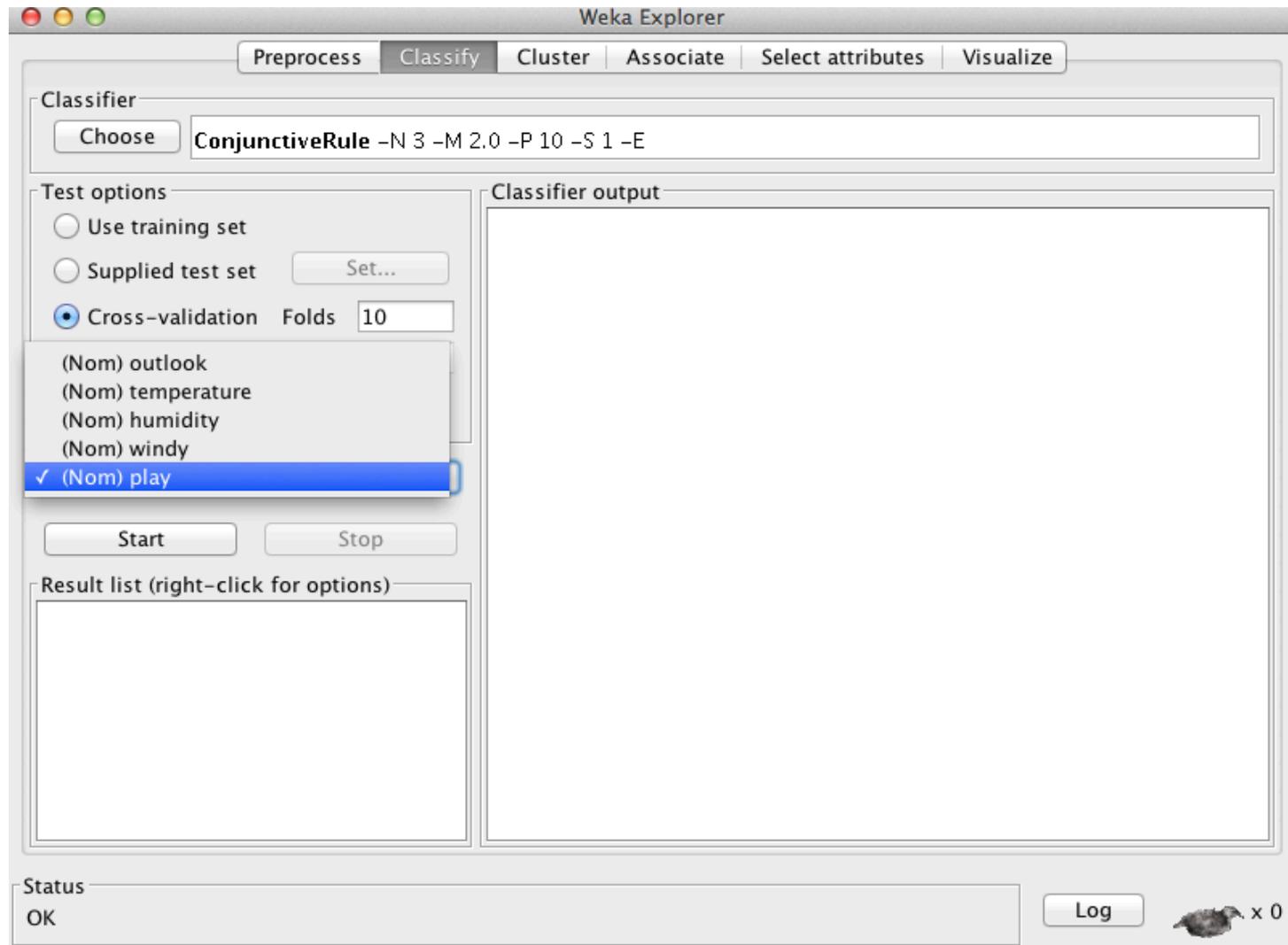
Result list (right-click for options)

Classifier output

Status OK   x 0

Even if you do not understand for now, select “Cross validation” with 10 Folds.

# Select the right hand side of the rule (the classification function)



# Run the algorithm

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'ConjunctiveRule -N 3 -M 2.0 -P 10 -S 1 -E'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10. The 'Start' button is highlighted with an orange arrow. The 'Classifier output' pane shows the following text:

```
Relation: weather.symbolic
Instances: 14
Attributes: 5
          outlook
          temperature
          humidity
          windy
          play
Test mode:10-fold cross-validation

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

Single conjunctive rule learner:
-----
(outlook = overcast) => play = yes

Class distributions:
Covered by the rule:
yes    no
1      0

Not covered by the rule:
yes    no
```

The 'Result list' shows '12:28:13 - rules.ConjunctiveRule' selected. The status bar at the bottom indicates 'OK' and a 'Log' button is visible.

# Performance data

The screenshot shows a software interface for a classifier. The main window is titled "Classifier" and contains several sections:

- Classifier:** A "Choose" button is followed by the text "ConjunctiveRule -N 3 -M 2.0 -P 10 -S 1 -E".
- Test options:** This section contains four radio buttons: "Use training set", "Supplied test set" (with a "Set..." button), "Cross-validation" (which is selected), and "Percentage split". The "Cross-validation" option has a "Folds" input field with the value "10" and a "Percentage split" input field with the value "66". A "More options..." button is located below these options.
- Classifier output:** This section displays the results of the classifier. It starts with "Single conjunctive rule learner:" followed by a dashed line and the rule "(outlook = overcast) => play = yes". Below this, it shows "Class distributions:" and "Covered by the rule:" with a table of counts for "yes" and "no" classes. It also shows "Not covered by the rule:" with a table of counts. The time taken to build the model is "0 seconds".
- Result list (right-click for options):** This section shows a list of results, with the first entry being "12:28:13 - rules.ConjunctiveRule".

The "Classifier output" section contains the following text:

```
Single conjunctive rule learner:
-----
(outlook = overcast) => play = yes

Class distributions:
Covered by the rule:
yes    no
1      0

Not covered by the rule:
yes    no
0.5    0.5

Time taken to build model: 0 seconds

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

Correctly Classified Instances      5      35.7143 %
Incorrectly Classified Instances    9      64.2857 %
Kappa statistic                    -0.4651
Mean absolute error                 0.4206
Root mean squared error             0.5113
Relative absolute error             88.3333 %
Root relative squared error        103.6427 %
```

# Error measures in WEKA

Mean absolute error is:

$$MSE = \frac{1}{N} \sum_{i=1}^N |\hat{\theta}_i - \theta_i|$$

Root mean square error is:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}$$

Where  $\theta_i$  is the “true” classification for  $x_i$  and  $\hat{\theta}_i$  is the classification produced by the ML algorithm (for binary classifiers  $\theta$  is either 0 or 1 and  $|\hat{\theta}_i - \theta_i|$  is either 0 or 1).

$$\sum_{i=1}^N |\hat{\theta}_i - \theta_i|$$

where  $\bar{\theta}$  is a mean value of  $\theta$ .

Root relative squared error:

$$RRSE = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}{\sum_{i=1}^N (\bar{\theta} - \theta_i)^2}}$$

# Confusion Matrix

	System classified as a	System classified as b
Truly classified as a	# of instances that system classifies a, ground truth is a	# of instances that system classifies b, ground truth is a
Truly classified as b	# of instances that system classifies a, ground truth is b	# of instances that system classifies b, ground truth is b

=== Confusion Matrix ===

```
a b  <-- classified as
5 4  | a = yes
5 0  | b = no
```

Cells (1,1) and (2,2) represent “good” classifications. The others are wrong. In fact, we are told that there are 5 correctly classified instances and 9 errors.