Machine Learning with WEKA

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

http://www.cs.waikato.ac.nz/ml/weka/
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 3.2: “GUI version” adds graphical user interfaces (book version is command-line only)
  - WEKA 3.3: “development version” with lots of improvements
  - .........................

- Latest stable version is WEKA 3.8
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
...

http://www.cs.waikato.ac.nz/ml/weka/arff.html
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
...
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

Current relation
Relation: iris
Instances: 150
Attributes: 5

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

Attributes
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sepal_length</td>
</tr>
<tr>
<td>2</td>
<td>sepal_width</td>
</tr>
<tr>
<td>3</td>
<td>petal_length</td>
</tr>
<tr>
<td>4</td>
<td>petal_width</td>
</tr>
<tr>
<td>5</td>
<td>class</td>
</tr>
</tbody>
</table>

Statistic          Value
Minimum             4.3
Maximum             7.9
Mean                5.843
StdDev              0.828

Colour: class (Nom)
Weka Knowledge Explorer

Filter
Choose: None

Current relation
Relation: iris
Instances: 150 Attributes: 5

Selected attribute
Name: petallength
Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)
Type: Numeric

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

Attributes
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sepalwidth</td>
</tr>
<tr>
<td>2</td>
<td>sepalwidth</td>
</tr>
<tr>
<td>3</td>
<td>petallength</td>
</tr>
<tr>
<td>4</td>
<td>petallwidth</td>
</tr>
<tr>
<td>5</td>
<td>class</td>
</tr>
</tbody>
</table>

Colour: class (Nom)
Visualize All

Status
OK

Log
x 0
Filter
Choose Discretize -B 10 -R first-last

Current relation
Relation: iris
Instances: 150
Attributes: 5

Attributes
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sepalwidth</td>
</tr>
<tr>
<td>2</td>
<td>sepalwidth</td>
</tr>
<tr>
<td>3</td>
<td>petallength</td>
</tr>
<tr>
<td>4</td>
<td>petalwidth</td>
</tr>
<tr>
<td>5</td>
<td>class</td>
</tr>
</tbody>
</table>

Selected attribute
Name: petallength
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
Choose Discretize –B 10 –R first–last
weka.filters.unsupervised.attribute.Discretize

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

Attributes:
- attributeIndices: first–last
- bins: 10
- findNumBins: False
- invertSelection: False
- makeBinary: False
- useEqualFrequency: False

Visualize All

Status
OK
### Current relation

**Relation:** iris  
**Instances:** 150

### Attributes

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>attributeIndices</th>
<th>bins</th>
<th>findNumBins</th>
<th>invertSelection</th>
<th>makeBinary</th>
<th>useEqualFrequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sepal length</td>
<td>first–last</td>
<td></td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>2</td>
<td>sepal width</td>
<td></td>
<td></td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>3</td>
<td>petal length</td>
<td>first–last</td>
<td></td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>4</td>
<td>petal width</td>
<td></td>
<td></td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>5</td>
<td>class</td>
<td></td>
<td></td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
</tbody>
</table>

### Visualize

[Histogram of Attribute Distribution]

**Status:** OK
Filter

Choose Discretize -F -B 10 -R first-last

Current relation
Relation: iris
Instances: 150  Attributes: 5

Attributes
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sepal length</td>
</tr>
<tr>
<td>2</td>
<td>sepal width</td>
</tr>
<tr>
<td>3</td>
<td>petal length</td>
</tr>
<tr>
<td>4</td>
<td>petal width</td>
</tr>
<tr>
<td>5</td>
<td>class</td>
</tr>
</tbody>
</table>

Selected attribute
Name: petal length
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
In the Weka Knowledge Explorer, the Discretize filter is applied to the dataset. The selected attribute is petallength, which is numeric. The statistics for petallength include:

- Minimum: 1
- Maximum: 6.9
- Mean: 3.759
- StdDev: 1.764

The dataset contains 150 instances and 5 attributes.
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, …
Choose ZeroR

Test options
- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

Classifier output

Result list (right-click for options)

Status: OK
Choose J48 -C 0.25 -M 2
Weka Knowledge Explorer

Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test 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
- 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
Weka Knowledge Explorer

Classifier
Choose J48 -C 0.25 -M 2

Test options
- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

Classifier output

### Run information ###
Scheme: weka.classifiers.trees.j48.J48 -C 0.25 -M 2
Relation: iris
Instances: 150
Attributes: 5
- sepal length
- sepal width
- petal length
- petal width
- class
Test mode: split 66% train, remainder test

### Classifier model (full training set) ###

J48 pruned tree

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

Number of Leaves : 5
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:

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Iris-versicolor</td>
<td>0.882</td>
<td>0.063</td>
<td>0.905</td>
<td>1</td>
<td>0.95</td>
</tr>
<tr>
<td>Iris-virginica</td>
<td>0.882</td>
<td>0.063</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
</tr>
</tbody>
</table>

Confusion Matrix:

```
<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>--- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>15</td>
<td>0</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>
```
Classifier:

Choose: J48 -C 0.25 -M 2

Test options:
- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

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:

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>0.882</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.882</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

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
Weka Knowledge Explorer

Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set
- Cross-validation
- Percentage split % 66

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 ===

<table>
<thead>
<tr>
<th>Class</th>
<th>Recall</th>
<th>F-Measure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Iris-versicolor</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

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

Choose J48 -C 0.25 -M 2

Test options
- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

Summary

<table>
<thead>
<tr>
<th>Correctly Classified Instances</th>
<th>Incorrectly Classified Instances</th>
<th>Kappa statistic</th>
<th>Mean absolute error</th>
<th>Root mean squared error</th>
<th>Relative absolute error</th>
<th>Root relative squared error</th>
<th>Total Number of Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>2</td>
<td>0.9408</td>
<td>0.0396</td>
<td>0.1579</td>
<td>8.8979</td>
<td>33.4091</td>
<td>51</td>
</tr>
</tbody>
</table>

Detailed Accuracy By Class

<table>
<thead>
<tr>
<th>Class</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-versicolor</td>
<td>1</td>
<td>0.95</td>
</tr>
<tr>
<td>Iris-virginica</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 ---
<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>1</td>
<td>0.063</td>
<td>0.905</td>
<td>1</td>
<td>0.95</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

--- Confusion Matrix ---

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>--- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>a = Iris-setosa</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>0</td>
<td>b = Iris-versicolor</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>15</td>
<td>c = Iris-virginica</td>
</tr>
</tbody>
</table>
Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

(Class) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48j48

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 ===

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>1</td>
<td>0.063</td>
<td>0.905</td>
<td>1</td>
<td>0.95</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

=== Confusion Matrix ===

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

Summary ===

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Correctly Classified Instances</td>
<td>49</td>
<td>96.0784 %</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
<td>2</td>
<td>3.9216 %</td>
</tr>
<tr>
<td>Statistic</td>
<td>0.9408</td>
<td></td>
</tr>
<tr>
<td>Mean squared error</td>
<td>0.1579</td>
<td></td>
</tr>
<tr>
<td>Relative squared error</td>
<td>33.4091 %</td>
<td></td>
</tr>
<tr>
<td>Number of Instances</td>
<td>51</td>
<td></td>
</tr>
</tbody>
</table>

Detailed Accuracy By Class ===

<table>
<thead>
<tr>
<th></th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>Iris-versicolor</td>
<td>0.063</td>
<td>0.905</td>
<td>1</td>
<td>0.95</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>Iris-virginica</td>
<td>0</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

Confusion Matrix ===

```
<table>
<thead>
<tr>
<th>0</th>
<th>15</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>
```

<-- classified as

```
a = Iris-setosa
b = Iris-versicolor
c = Iris-virginica
```
Classifier:

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
- 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 0.0396
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

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>1</td>
<td>0.063</td>
<td>0.905</td>
<td>1</td>
<td>0.95</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

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

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

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

a b c  <- classified as
15 0 0  a = Iris-setosa
0 19 0  b = Iris-versicolor
0 2 15  c = Iris-virginica
### 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**
- **Cross-validation**
- **Percentage split** % 66

#### Classifier output

- **Evaluation on test split**

  **Summary**

  | Correctly Classified Instances | 50 |
  | Incorrectly Classified Instances | 1 |
  | 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**

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>1</td>
<td>0.031</td>
<td>0.95</td>
<td>1</td>
<td>0.974</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.941</td>
<td>0</td>
<td>1</td>
<td>0.941</td>
<td>0.97</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

- **Confusion Matrix**

  
<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>---</th>
<th>classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>a</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>0</td>
<td>b</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>16</td>
<td>c</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

### Status

**OK**
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 ===

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>Iris-virginica</td>
<td>0.941</td>
<td>0.031</td>
<td>0.95</td>
<td>1</td>
<td>0.974</td>
<td>Iris-virginica</td>
</tr>
<tr>
<td>Iris-versicolor</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.941</td>
<td>0.97</td>
<td>Iris-versicolor</td>
</tr>
</tbody>
</table>

Confusion Matrix ===

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
</tbody>
</table>

classified as
a = Iris-setosa
b = Iris-versicolor
c = Iris-virginica
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
NaiveBayes

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:

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>0.947</td>
<td>0.063</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0.029</td>
<td>0.938</td>
<td>0.882</td>
<td>0.909</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

Confusion Matrix:

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>--- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>a = Iris-setosa</td>
</tr>
<tr>
<td>0</td>
<td>18</td>
<td>1</td>
<td>b = Iris-versicolor</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>15</td>
<td>c = Iris-virginica</td>
</tr>
</tbody>
</table>
### 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
- 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 ---

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>Iris-versicolor</td>
<td>0.938</td>
<td>0.947</td>
<td>0.923</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>Iris-virginica</td>
<td>0.938</td>
<td>0.882</td>
<td>0.909</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

--- 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
Weka Knowledge Explorer

Classifier

Choose NaiveBayes

Test options

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

More options

(Nom) class

Start

Result list (right-click for option)

- 11:49:05 - trees.j48,J48
- 14:34:28 - functions.neu
- 14:48:05 - bayes.NaiveBayes

Plot: ThresholdCurve

Class colour

Status

OK
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

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 ---

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris-setosa</td>
<td>1.00</td>
<td>0.063</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Iris-versicolor</td>
<td>0.947</td>
<td>0.063</td>
<td>0.90</td>
<td>0.947</td>
<td>0.923</td>
</tr>
<tr>
<td>Iris-virginica</td>
<td>0.882</td>
<td>0.029</td>
<td>0.938</td>
<td>0.882</td>
<td>0.909</td>
</tr>
</tbody>
</table>

--- Confusion Matrix ---

```
15 0 0  <!-- classified as
0 18 1  a = Iris-setosa
0 2 15  b = Iris-versicolor
```
== Evaluation on test split ==
== Summary ==

correctly Classified Instances 48 94.1176 %
correctly Classified Instances 3 5.8824 %

apparent statistic 0.9113
ean absolute error 0.0447
oot mean squared error 0.1722
e relative absolute error 10.0365 %
e relative squared error 36.4196 %
total Number of Instances 51

== Detailed Accuracy By Class ==

<table>
<thead>
<tr>
<th>P Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>0.947</td>
<td>0.063</td>
<td>0.9</td>
<td>1</td>
<td>0.947</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0.029</td>
<td>0.938</td>
<td>0.882</td>
<td>0.909</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

== Confusion Matrix ==

a  b  c
15 0 0 | a = Iris-setosa
0 18 1 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Classifier
Choose UserClassification

Test options
- Use training set
- Supplied test set
- Cross-validation
- Percentage split

(Nom) class

Start

Result list (right-click for options)
14:34:28 - functions.neural
14:48:05 - bayes.NaiveBayes
15:26:57 - trees.UserClassification

Status
Building model on training data...

Tree View
- Iris-setosa, 50.0
- Iris-versicolor, 50.0
- Iris-virginica, 50.0
Building model on training data...
Split on petallength AND petalwidth

True
- Iris-versicolor, 1.0
- Iris-virginica, 48.0

False
- Iris-setosa, 50.0
- Iris-versicolor, 49.0
- Iris-virginica, 2.0
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

--- 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 ---
<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>1</td>
<td>0.063</td>
<td>0.905</td>
<td>1</td>
<td>0.95</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

--- 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
Weka Knowledge Explorer

Classifier

Choose: UserClassifier

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

More options

(Num) sepalwidth
(Num) petalwidth
(Num) petallength
(Nom) class

Class: (Num) petallength

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
Mean squared error 0.1622
Relative absolute error 7.1634%
Root mean squared error 34.312%
Total Number of Instances 51

--- Detailed Accuracy By Class ---

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Iris-setosa</td>
</tr>
<tr>
<td>1</td>
<td>0.063</td>
<td>0.905</td>
<td>0.95</td>
<td>0.95</td>
<td>Iris-versicolor</td>
</tr>
<tr>
<td>0.882</td>
<td>0</td>
<td>1</td>
<td>0.882</td>
<td>0.938</td>
<td>Iris-virginica</td>
</tr>
</tbody>
</table>

--- 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.
M5 pruned model tree:
(using smoothed predictions)

petalwidth <= 0.8 : LM1 (50/10.469%)
petalwidth > 0.8 :
  | class=Iris-virginica <= 0.5 : LM2 (50/14.325%)
  | class=Iris-virginica > 0.5 : LM3 (50/17.598%)

LM num: 1
Linear Regression Model

petallength =
  0.4652 + petalwidth
Classifier output

petallength =  
\[
0.4208 \times \text{sepal length} + 1.2692 \times \text{petal width} + 0.0795
\]

LM num: 3
Linear Regression Model

petallength =  
\[
0.7501 \times \text{sepal length} + 0.6105
\]

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
M5P Classifier

Test options:
- Use training set
- Supplied test set: Set...
- Cross-validation: Folds 10
- Percentage split: % 66

(Num) petallength

Classification:
- Linear Regression Model
  - petallength = 0.7501 * sepalwidth + 0.6105

Number of Rules: 3
Time taken to build model: 1.31 seconds

Evaluation on test split:
- 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
Weka Knowledge Explorer

Classifier: M5P

Test options:
- X: sepal length (Num)
- Y: petal width (Num)
- Colour: petal length (Num)

Plot: iris_predicted

Class colour:
- 1.1
- 3.75
- 6.4

Root relative squared error: 14.9153%
Total Number of Instances: 51
Instance 31:

- Instance_number: 31.0
- sepal_length: 6.9
- sepal_width: 3.1
- predicted_petal_length: 5.892812341943582
- petal_length: 5.1
- petal_width: 2.3
- class: Iris-virginica
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
Choose EM -l 100 -N 1 -S 100 -M 1.0E-6

Cluster mode
- Use training set
- Supplied test set
- Percentage split
- Classes to clusters evaluation
  (Nom) class
- Store clusters for visualization

Ignore attributes

Start
Stop

Clusterer output

Result list (right-click for options)

Status
OK
Clusterer

Choose: Cobweb - A 1.0 - C 0.0028209479177387815

Cluster mode

- Use training set
- Supplied test set
- Percentage split
- Classes to clusters evaluation

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]
  | leaf 2 [54]

Clustered Instances

1 100 (67%)
2 50 (33%)

Class attribute: class
Classes to Clusters:

1 2 <-- assigned to cluster
0 50 | Iris-setosa
50 0 | Iris-versicolor
50 0 | Iris-virginica

Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances: 50.0 33.3333 %
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

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 <-- assigned to cluster
0 50  Iris-setosa
50 0  Iris-versicolor
50 0  Iris-virginica

Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances: 50.0 33.33333 %
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

Status
OK
Clusterer

Choose: Cobweb - A 1.0 - C 0.0028209479177387815

Cluster mode

- Use training set
- Supplied test set
- Percentage split
- Classes to cluster

(Nom) class

Store clusters for future runs

Start

Result list (right-click for options)

16:05:58 - Cobweb

Clusterer output

Weka Classifier Tree Visualizer: 16:05:58 - Cobweb (iris)

Tree View

node 0 (150)

leaf 1 (96)

leaf 2 (54)

Status

OK
Weka Knowledge Explorer

Clusterer
Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode
- Use training set
- Supplied test 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
  sepal length
  sepal width
  petal length
  petal width
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
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
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
Weka Knowledge Explorer

Associate

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

Status
OK
Minimum metric <confidence>: 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
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 == Class=democrat 210
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y education-spending=n 202 == Class=democrat 201 conf:0.99
4. physician-fee-freeze=n 247 == Class=democrat 245 conf:0.98
5. el-salvador-aid=n Class=democrat 200 == aid-to-nicaraguan-contras=y 197 conf:0.99
6. el-salvador-aid=n 208 == aid-to-nicaraguan-contras=y 204 conf:0.98
7. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 218
8. aid-to-nicaraguan-contras=y 218 == physician-fee-freeze=n 210
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
Weka Knowledge Explorer

Attribute Evaluator
- Choose CfsSubsetEval

Search Method
- Choose BestFirst -D 1 -N 5

Attribute Selection Mode
- Use full training set

Attribute Selection Output

Start

Result list (right-click for options)

Status
OK
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
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

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

Selected attributes: 4 : 1
physician-fee-freeze
Attribute evaluator: InfoGainAttributeEval

Search method:
- BestFirst
- ForwardSelection
- RaceSearch
- GeneticSearch
- RandomSearch
- ExhaustiveSearch
- Ranker
- RankSearch

Attribute selection output:

<table>
<thead>
<tr>
<th>Duty-free-exports</th>
</tr>
</thead>
<tbody>
<tr>
<td>export-administration-act-south-africa</td>
</tr>
<tr>
<td>Class</td>
</tr>
</tbody>
</table>

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
Choose InfoGainAttributeEval

Choose Ranker -T -1.7976931348623157E308 -N -1

Use full training set

Folds 10
Seed 1

(Nom) Class

Start Stop

Result list (right-click for options)

16:39:40 - BestFirst + CfsSubsetEval
16:43:05 - Ranker + InfoGainAttributeEval

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-missle
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
0.0490977  10 immigration
0.0049097  16 selected attributes: 4,3,5,12,14,8,9,13,15,7,6,1,11,16,10,2 : 16
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
Relation: Glass
Instances: 214
Attributes: 10

Selected attribute:
Name: RI
Missing: 0 (0%)
Distinct: 178
Unique: 145 (68%)

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

Colour: Type (Nom)
Weka Knowledge Explorer: Visualizing Glass

- X: Al (Num)
- Y: Ca (Num)
- Colour: Type (Nom)
- Rectangle

Plot: Glass

Class colour:
- build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps
Conclusion: try it yourself!

- WEKA is available at
  http://www.cs.waikato.ac.nz/ml/weka

- Slides are adopted from
  - Department of Computer Science,
    University of Waikato, New Zealand