WEKA

Introduction
WEKA software

- WEKA homepage
  - http://www.cs.waikato.ac.nz/ml/weka/
WEKA software

- Data mining software written in Java (distributed under the GNU Public License).
- Used for research, education, and applications.
- Comprehensive set of data preprocessing tools, learning algorithms and evaluation methods.
- Graphical user interfaces – including data visualization.
- Environment for comparing learning algorithms.
- Book and development versions.
ARFF format file

@relation data-relation-name

@attribute attribute1 real

@attribute attribute2 { nominal1, nominal2 }

@attribute class { class1, class2 }

@data
1, nominal1, class1
2, nominal2, class2
3, ?, class1
4, nominal1, class2
The Weka GUI Chooser (class weka.gui.GUIChooser) provides a starting point for launching Weka’s main GUI applications and supporting tools. If one prefers a MDI ("multiple document interface") appearance, then this is provided by an alternative launcher called “Main” (class weka.gui.Main).

The GUI Chooser consists of four buttons—one for each of the four major Weka applications—and four menus.
WEKA GUI Chooser

Weka GUI Chooser

Applications

Explorer

Experimenter

KnowledgeFlow

Simple CLI

Waikato Environment for Knowledge Analysis
Version 3.6.9
(c) 1999 - 2013
The University of Waikato
Hamilton, New Zealand
WEKA Experimenter

- An environment for performing experiments and conducting statistical tests between learning schemes.
- The Weka Experiment Environment enables the user to create, run, modify, and analyse experiments in a more convenient manner than is possible when processing the schemes individually. For example, the user can create an experiment that runs several schemes against a series of datasets and then analyse the results to determine if one of the schemes is (statistically) better than the other schemes.
WEKA Experimenter
This environment supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning.

The KnowledgeFlow presents a data-flow inspired interface to WEKA. The user can select WEKA components from a tool bar, place them on a layout canvas and connect them together in order to form a knowledge flow for processing and analyzing data. At present, all of WEKA’s classifiers, filters, clusterers, loaders and savers are available in the KnowledgeFlow along with some extra tools.
WEKA Simple CLI

- Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface.
- The Simple CLI provides full access to all Weka classes, i.e., classifiers, filters, clusterers, etc., but without the hassle of the CLASSPATH (it facilitates the one, with which Weka was started).
- It offers a simple Weka shell with separated commandline and output.
WEKA Simple CLI

Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of the window. Use the up and down arrows to move through previous commands.
Command completion for classnames and files is initiated with <Tab>. In order to distinguish between files and classnames, file names must be either absolute or start with './' or '/'
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text in the commandline in chunks.

> help

Command must be one of:
java <classname> <args> [ > file]
break
kill
cls
history
exit
help <command>
WEKA Explorer

- An environment for exploring data with WEKA (the rest of this documentation deals with this application in more detail).
- At the very top of the window, just below the title bar, is a row of tabs. When the Explorer is first started only the first tab is active; the others are greyed out. This is because it is necessary to open (and potentially pre-process) a data set before starting to explore the data.
WEKA Explorer
WEKA Explorer

- The tabs are as follows:
  - **Preprocess**: Choose and modify the data being acted on.
  - **Classify**: Train and test learning schemes that classify or perform regression.
  - **Cluster**: Learn clusters for the data.
  - **Associate**: Learn association rules for the data.
  - **Select attributes**: Select the most relevant attributes in the data.
  - **Visualize**: View an interactive 2D plot of the data.
Selecting a Classifier
At the top of the classify section is the Classifier box. This box has a text field that gives the name of the currently selected classifier, and its options. The Choose button allows you to choose one of the classifiers that are available in WEKA.

There are four test modes available:

- **Use training set**: The classifier is evaluated on how well it predicts the class of the instances it was trained on.
- **Supplied test set**: The classifier is evaluated on how well it predicts the class of a set of instances loaded from a file. Clicking the Set... Button brings up a dialog allowing you to choose the file to test on.
- **Cross-validation**: The classifier is evaluated by cross-validation, using the number of folds that are entered in the Folds text field.
- **Percentage split**: The classifier is evaluated on how well it predicts a certain percentage of the data which is held out for testing. The amount of data held out depends on the value entered in the % field.

Once the classifier, test options and class have all been set, the learning process is started by clicking on the Start button. While the classifier is busy being trained, the little bird moves around.
WEKA Explorer - Classification

Classifier output:

--- Summary ---

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
<td>3068</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
<td>265</td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.6639</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.095</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.265</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>38.3215 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>75.2804 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>3333</td>
</tr>
</tbody>
</table>

--- 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>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.671</td>
<td>0.037</td>
<td>0.753</td>
<td>0.671</td>
<td>0.71</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>0.963</td>
<td>0.329</td>
<td>0.945</td>
<td>0.960</td>
<td>0.954</td>
<td>0.029</td>
</tr>
<tr>
<td>Weighted Avg.</td>
<td>0.92</td>
<td>0.287</td>
<td>0.917</td>
<td>0.92</td>
<td>0.919</td>
<td>0.829</td>
</tr>
</tbody>
</table>

--- Confusion Matrix ---

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>324</td>
<td>159</td>
</tr>
<tr>
<td>106</td>
<td>2744</td>
</tr>
</tbody>
</table>

Status: OK
When training is complete, several things happen. The Classifier output area to the right of the display is filled with text describing the results of training and testing. A new entry appears in the Result list box.

The text in the Classifier output area is split into several sections:

- **Run information**: A list of information giving the learning scheme options, relation name, instances, attributes and test mode that were involved in the process.
- **Classifier model (full training set)**: A textual representation of the classification model that was produced on the full training data.
- **Summary**: A list of statistics summarizing how accurately the classifier was able to predict the true class of the instances under the chosen test mode.
- **Detailed Accuracy By Class**: A more detailed per-class break down of the classifier's prediction accuracy.
- **Confusion Matrix**: Shows how many instances have been assigned to each class. Elements show the number of test examples whose actual class is the row and whose predicted class is the column.
More can be found in WEKA Manual:

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