BUS2004

Week 2 Workshop

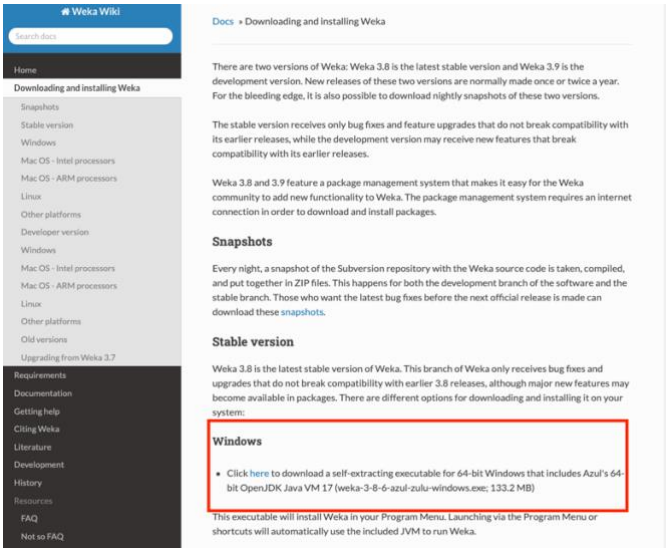
Introducing Weka for Machine Learning

1. Downloading and installing Weka

Weka can be downloaded and installed for most common operating systems. This workshop will look at installing Weka for Windows and macOS machines. Also, Weka can be installed in two ways: We can either download Weka with the required base Java libraries or we can download Weka on its own if, for example, we have all the required libraries installed beforehand. We will use the All-In-One installation method here. For the first step, download the version you need from http://www.cs.waikato.ac.nz/ml/weka/downloading.html

1.1. Weka for Windows

For Windows computers, the first step is to choose the 64-bit or the 32-bit version. You don't need to go into deep details. Most modern computers will require 64-bit. Therefore, please download the Self-extracting executable for 64-bit Windows that includes Oracle's 64-bit Java.



After the download, double-click the downloaded files and follow the prompts for installation. The Weka icon will appear in your start menu. Click it to start Weka.

1.2. Weka for MacOS

Following the same URL provided above, for MacOS we need to download Disk image for OS X that contains a Mac application including Oracle's Java.

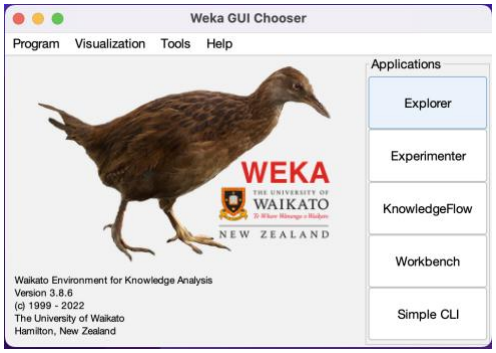
After extracting the downloaded folder, drag both the folder and the icon into your Applications folder.

After that, you can start Weka by clicking the bird icon.

2. Launching WEKA

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.



The buttons can be used to start the following applications:

- Explorer An environment for exploring data with WEKA.
- Experimenter An environment for performing experiments and conducting statistical tests between learning schemes.

- KnowledgeFlow 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.
- Workbench An all-in-one application that combines all the others within user-selectable “perspectives”.
- SimpleCLI 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 menus consist of:

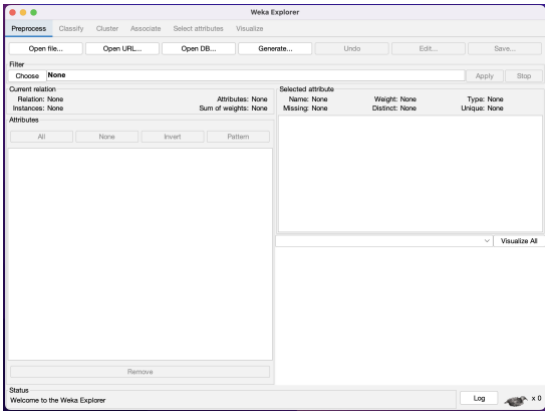
- Program
  - LogWindow Opens a log window that captures all that is printed to stdout or stderr. Useful for environments like MS Windows, where WEKA is normally not started from a terminal.
  - Exit Closes WEKA.
- Tools Other useful applications.
  - Package manager A graphical interface to Weka’s package management system.
  - ArffViewer An MDI application for viewing ARFF files in spreadsheet format.
  - SqlViewer Represents an SQL worksheet, for querying databases via JDBC.
  - Bayes net editor An application for editing, visualizing and learning Bayes nets.
- Visualization Ways of visualizing data with WEKA.
  - Plot For plotting a 2D plot of a dataset.
  - ROC Displays a previously saved ROC curve.
  - TreeVisualizer For displaying directed graphs, e.g., a decision tree.
  - GraphVisualizer Visualizes XML BIF or DOT format graphs, e.g., for Bayesian networks.
  - BoundaryVisualizer Allows the visualization of classifier decision boundaries in two dimensions.
- Help
  - Online resources for WEKA can be found here.
  - Weka homepage Opens a browser window with WEKA’s home- page.
  - HOWTOs, code snippets, etc

3. First Weka Example (The Explorer):

3.1. We're going to explore the Explorer, the Weka Explorer interface

3.2. There are four interfaces in Weka. The Explorer is the one that we'll be using throughout this class.

3.3. Across the top, there are five panels:



3.3.1. the Preprocess panel;

3.3.2. the Classify panel, where you build classifiers for datasets;

3.3.3. clustering, another procedure Weka is good at.

3.3.4. association rules.

3.3.5. attribute selection

3.3.6. visualization.

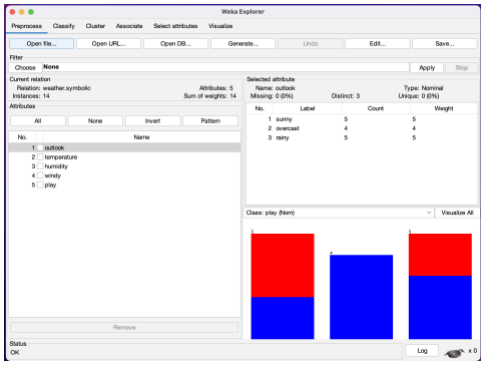
3.4. Let us open the sample weather dataset. This is a sample data of 14 days where for each day, we have 5 attributes four to do with the weather: Outlook, Temperature, Humidity, and Windy. The fifth, Play, is whether or not we're going to play some game. Think of this “play” as the results we will be predicting using the other 4 attributes.

3.5. To open the dataset, you need to locate the “data” folder inside your Weka folder.

3.5.1. For windows, try “C:\Program Files\Weka-3-8\data”

3.5.2. For Mac, try /Applications/weka-3.8.6.app/Contents/app/data or /usr/local/share/weka-3.8.6/data

3.6. Open weather.nominal.arff.



3.7. Select the attribute outlook to see the values. The values for the outlook attribute are sunny, overcast, and rainy. These are the number of times they appear in the dataset: 5 sunny days, 4 overcast days, and 3 rainy days for a total of 14 days.

3.8. If we look at the temperature attribute, hot, mild, and cool are the possible values, and these are the number of times they appear in the dataset.

3.9. Let's go to the play attribute. There are two values for play, yes or no.

3.10. Now, let's look at these two bars. Blue corresponds to yes, and red corresponds to no.

3.11. If you look at one of the other attributes, like outlook, you can see that when the outlook is sunny, there are three no instances and two yes instances.

3.12. Go to the Edit panel. This will display the data in the form that it is on the file. This is another view of the data.

3.13. You can edit this dataset. Click on any value in the table, then choose a different value from the drop-down list.

3.14. Click OK to see the change reflected on the graph.

3.15. The dataset on the disk is still the same as it was and the change has not been permanently saved.