

Numeric prediction in Weka

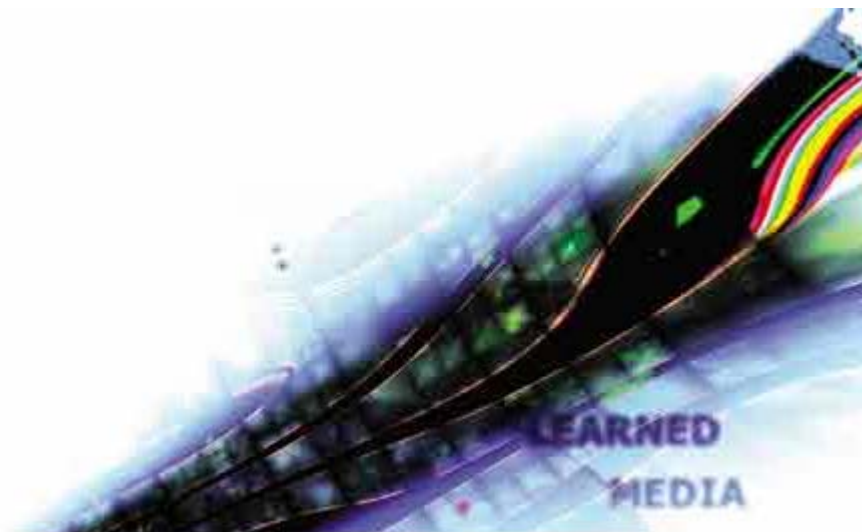
15.11.2007

Petra Kralj

Petra.Kralj@ijs.si

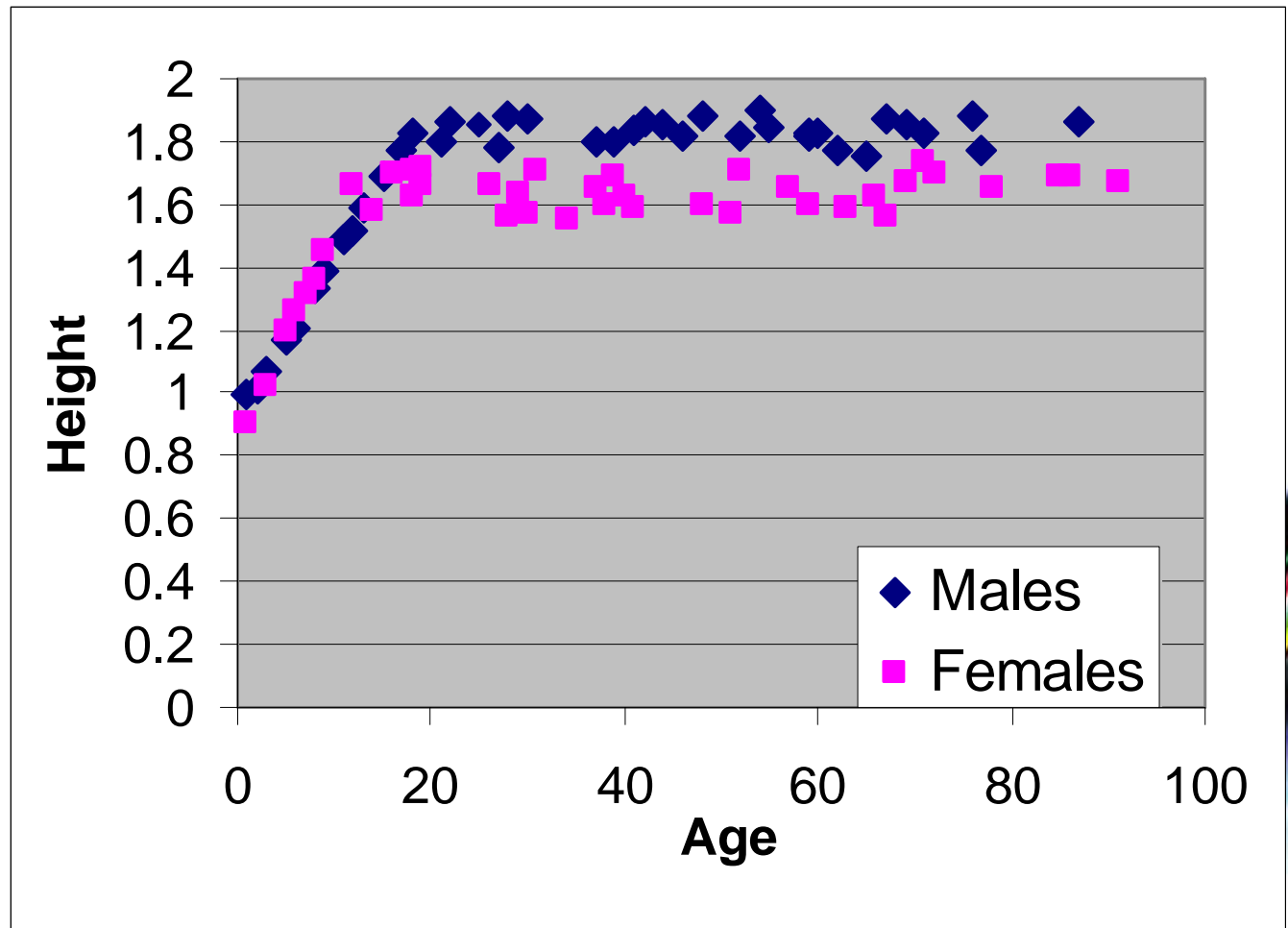
Numeric prediction models

- LinearRegression
- M5P Regression and model trees
- KNN
- Baseline predictor



regressionAgeHeight.csv

- Data about 80 people:
- Age
- Gender
- Height



Filename:

The screenshot shows the Weka Explorer application window. The title bar reads "Weka Explorer". The menu bar includes "Preprocess", "Classify", "Cluster", "Associate", "Select attributes", and "Visualize". Below the menu bar are buttons for "Open file...", "Open URL...", "Open DB...", "Undo", "Edit...", and "Save...".

The "Filter" section shows "Choose" and "None" buttons, with an "Apply" button to the right.

The "Current relation" section displays "Relation: regressionAheHeight" and "Instances: 80" with "Attributes: 3".

The "Attributes" section has "All", "None", and "Invert" buttons. Below is a table of attributes:

No.	Name
1	<input type="checkbox"/> Age
2	<input type="checkbox"/> Gender
3	<input checked="" type="checkbox"/> Height

Below the table is a "Remove" button.

The "Selected attribute" section shows "Name: Height", "Missing: 0 (0%)", "Distinct: 80", "Type: Numeric", and "Unique: 80 (100%)". Below this is a table of statistics:

Statistic	Value
Minimum	0.902
Maximum	1.895
Mean	1.628
StdDev	0.236

Below the statistics is a dropdown menu set to "Class: Height (Num)" and a "Visualize All" button.

The visualization area shows a histogram for the "Height" attribute. The x-axis ranges from 0.9 to 1.9. The y-axis represents frequency. The bars have the following heights: 5 (at 0.9), 4 (at 1.0), 6 (at 1.1), 30 (at 1.2), and 35 (at 1.3).

The "Status" bar at the bottom shows "OK". There is a "Log" button and a small icon with "x 0" next to it.



Visualization in Weka



Weka Explorer

Preprocess

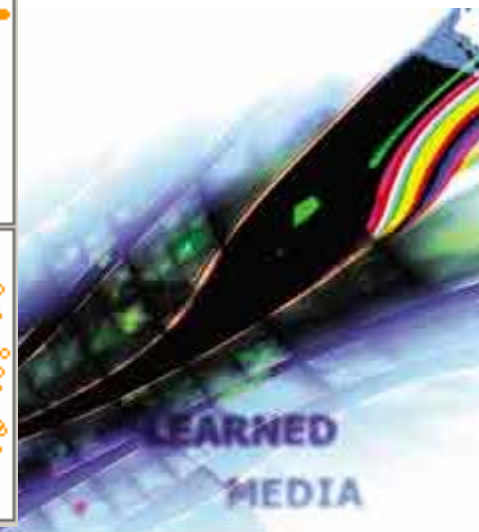
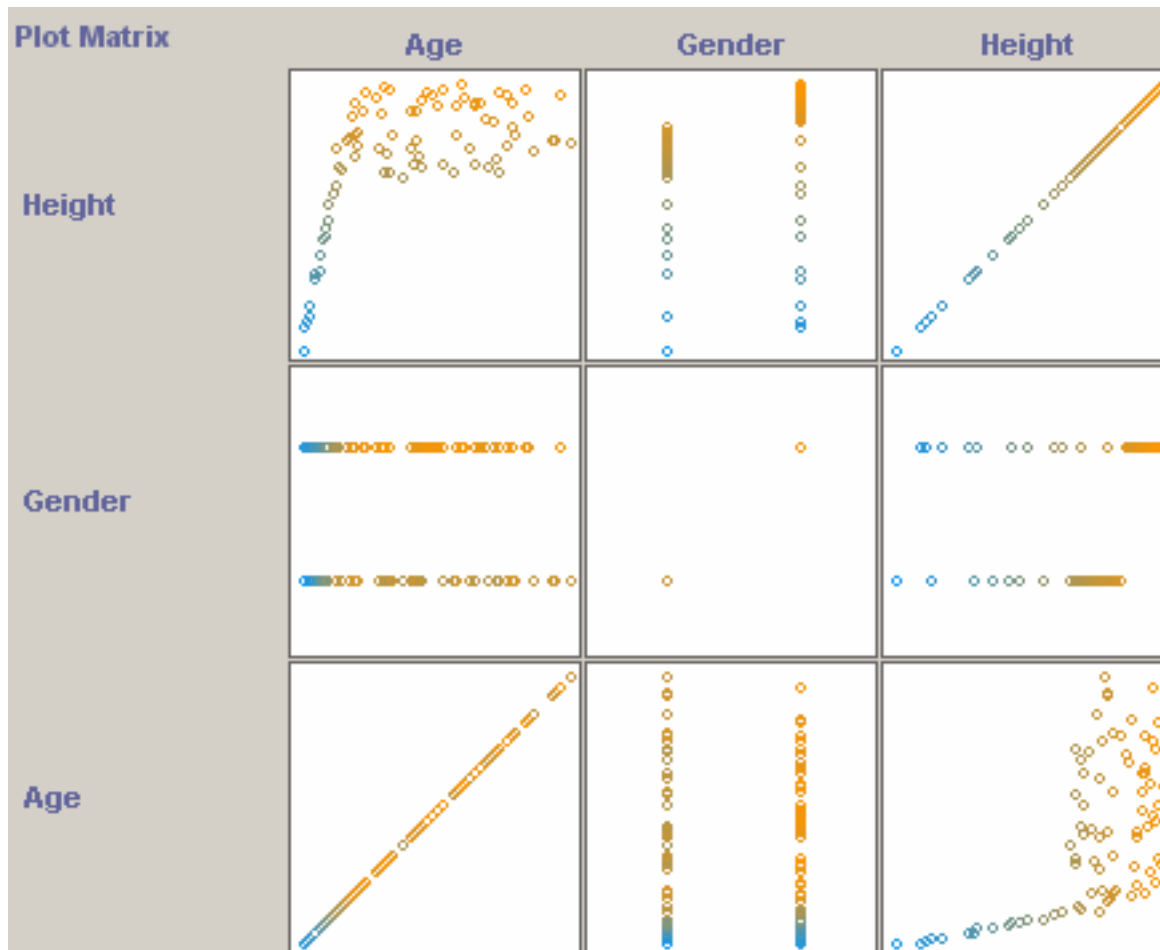
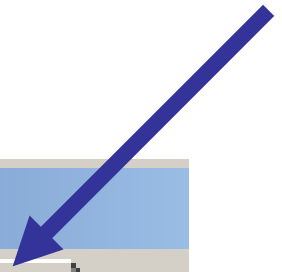
Classify

Cluster

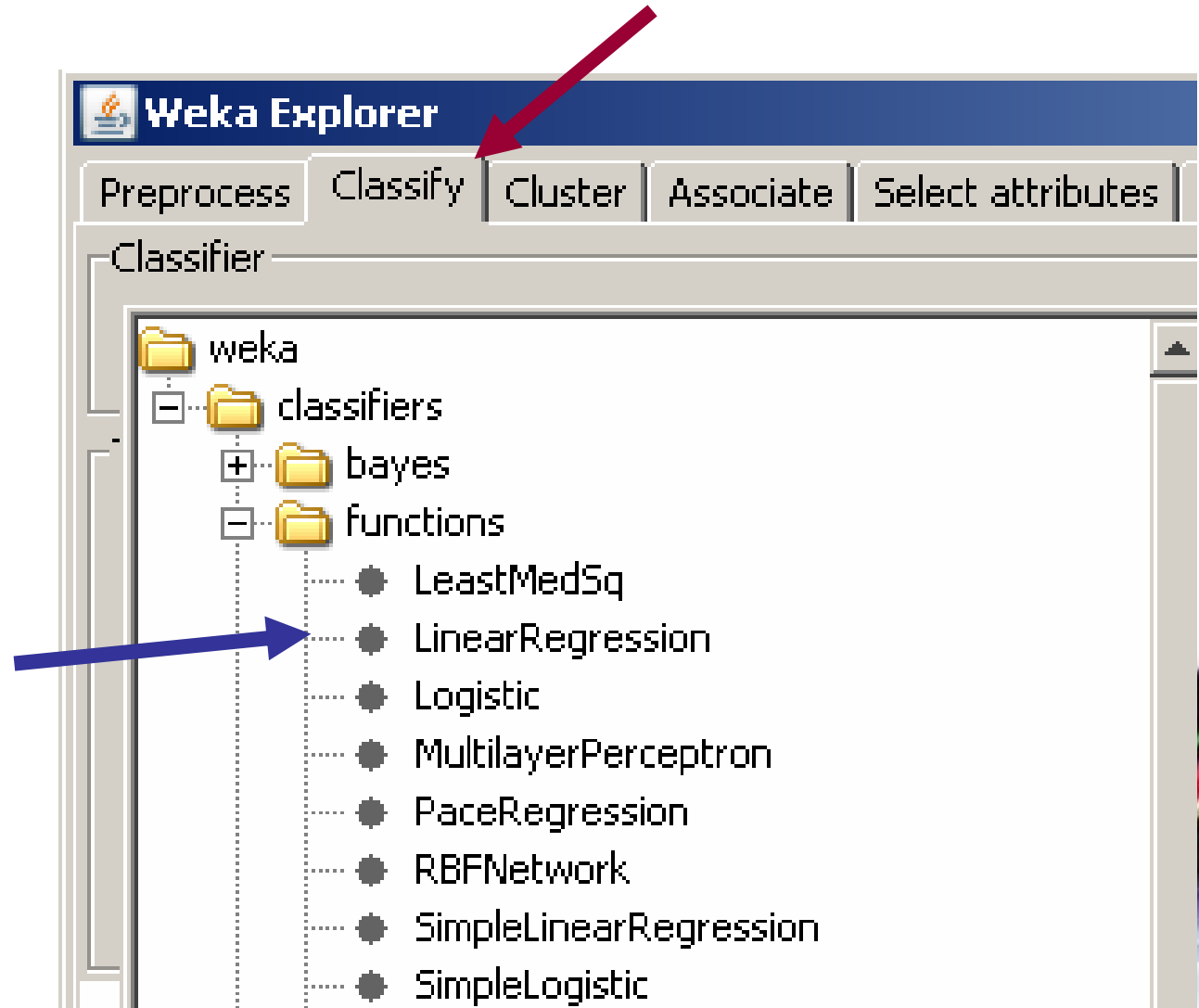
Associate

Select attributes

Visualize



Weka → classifiers → functions → LinearRegression



Weka Explorer

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

Classifier: Choose **MSP -M 4.0**

Test options:

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

(Num) Height

Result list (right-click for options)

- 15:24:52 - trees.M5P
- 15:58:13 - functions.LinearRegression**
- 16:03:32 - functions.LinearRegression

Classifier output:

```

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

Linear Regression Model


Height =

      0.0056 * Age +
      0.1292 * Gender=M +
      1.3506

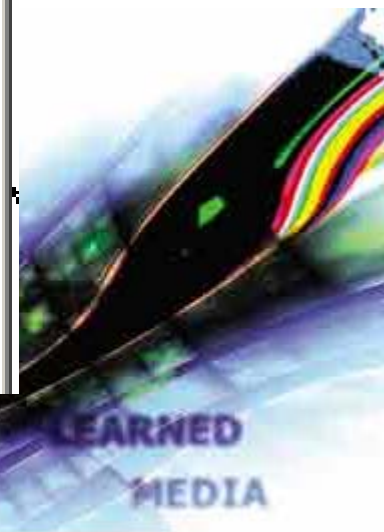
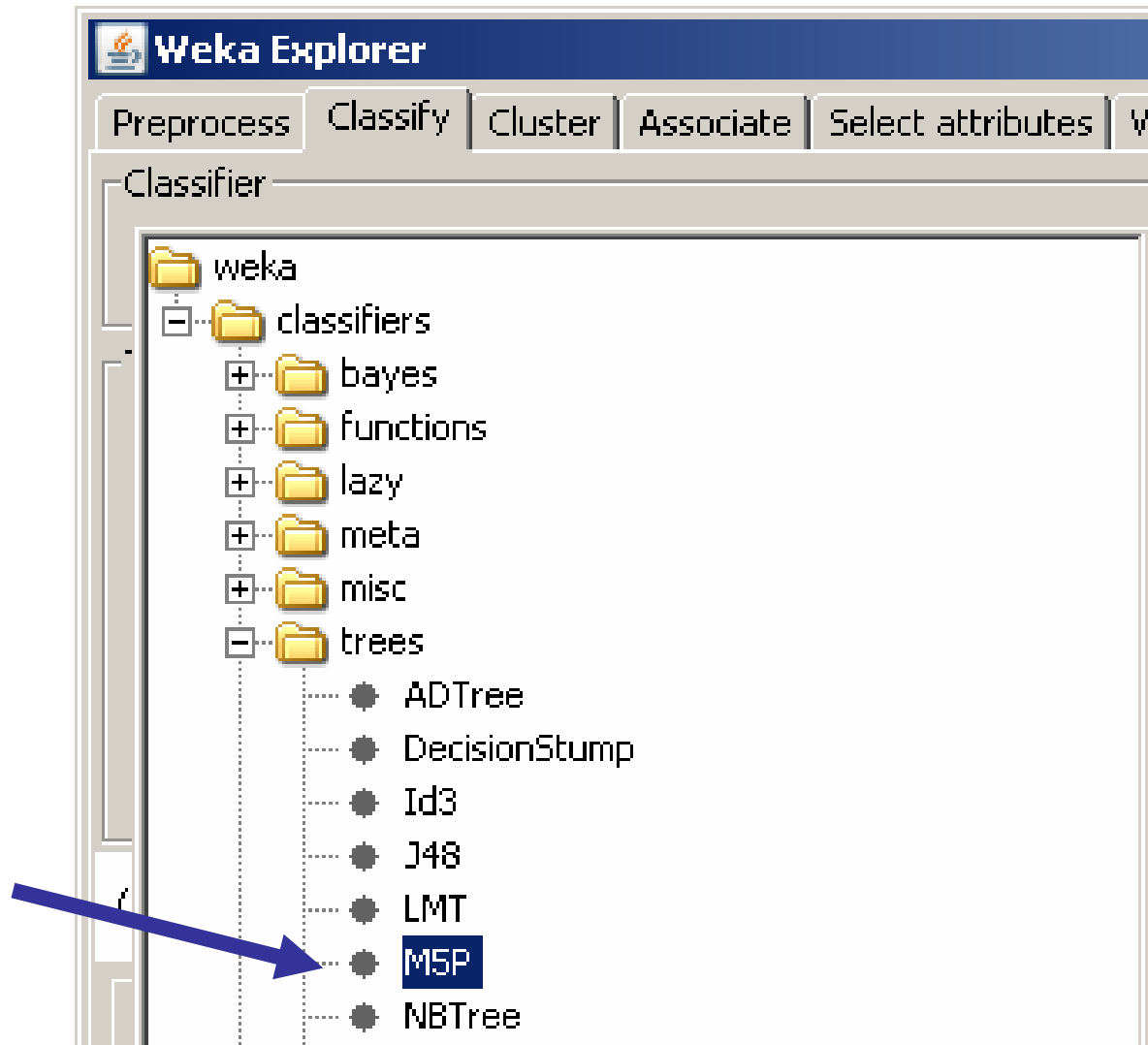
Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient           0.6204
Mean absolute error              0.142
Root mean squared error         0.1844
Relative absolute error         80.1623 %
Root relative squared error     77.2023 %
Total Number of Instances      80
  
```

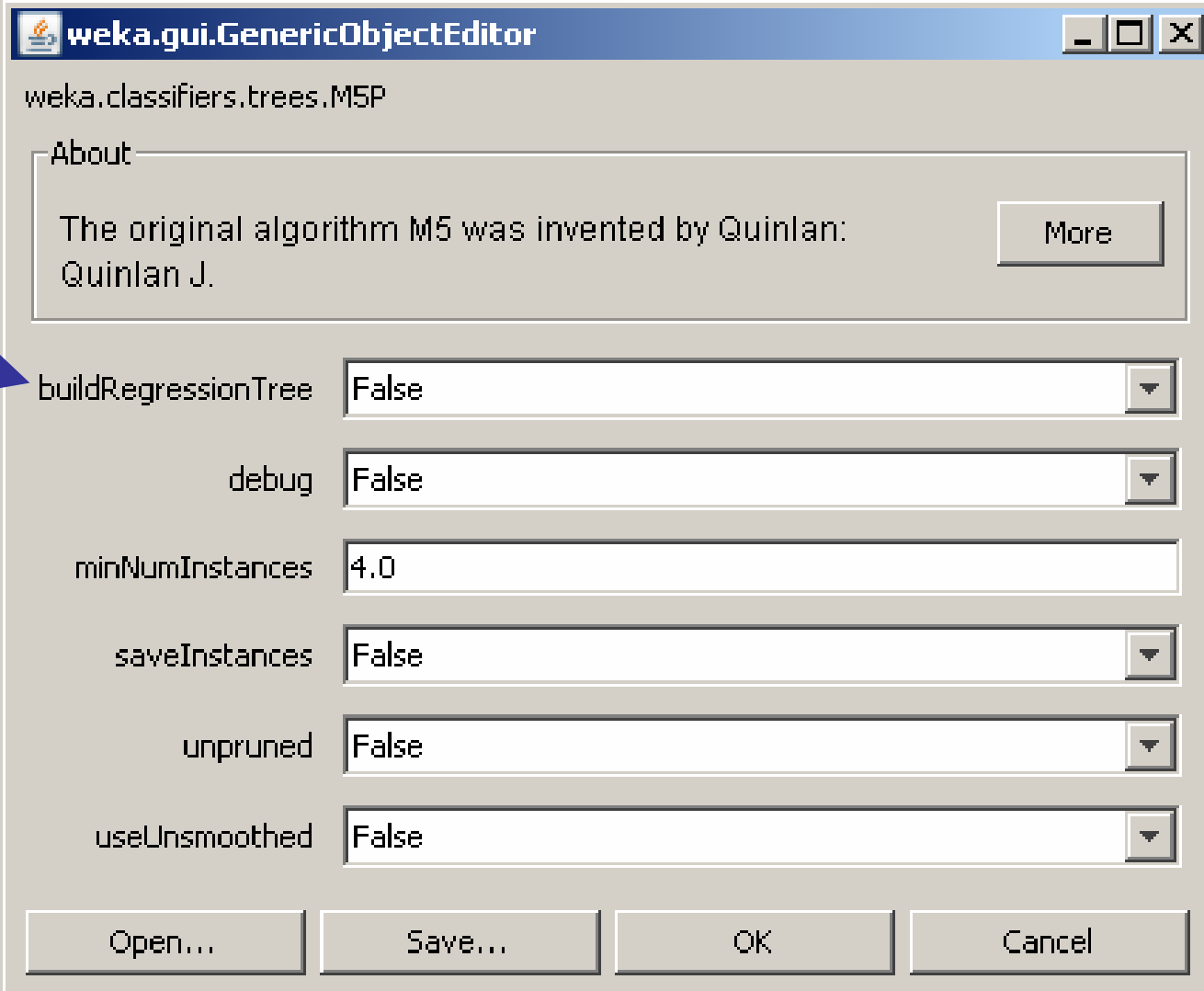
Status: OK  x 0

Weka → classifiers → trees → M5P



LEARNED
MEDIA

builtRegressionTree = True → regression tree
builtRegressionTree = False → model tree



The screenshot shows the 'weka.gui.GenericObjectEditor' window for the 'weka.classifiers.trees.M5P' classifier. The window contains an 'About' section with text: 'The original algorithm M5 was invented by Quinlan: Quinlan J.' and a 'More' button. Below this are several configuration options, each with a dropdown menu:

- builtRegressionTree: False
- debug: False
- minNumInstances: 4.0
- saveInstances: False
- unpruned: False
- useUnsmoothed: False

At the bottom of the window are four buttons: 'Open...', 'Save...', 'OK', and 'Cancel'. A blue arrow points to the 'builtRegressionTree' dropdown menu.

kNN: weka → classifiers → lazy
→ IBk

The image shows two windows from the Weka software. The left window, 'Weka Explorer', displays a tree view of the classifier hierarchy. A blue arrow points to the 'IBk' classifier under the 'lazy' folder. The right window, 'weka.gui.GenericObjectEditor', shows the configuration for the selected classifier. A blue arrow points from the 'IBk' classifier in the tree to the 'KNN' field in the editor, which is set to '3'. Other configuration options include 'crossValidate' (False), 'debug' (False), 'distanceWeighting' (No distance weighting), 'meanSquared' (False), 'noNormalization' (False), and 'windowSize' (0). The 'About' section describes it as a 'K-nearest neighbours classifier'.

Weka Explorer

Preprocess Classify Cluster

Classifier

- weka
 - classifiers
 - bayes
 - functions
 - lazy
 - IB1
 - IBk**
 - KStar
 - LBR
 - LWL
 - meta
 - misc
 - trees
 - rules

weka.gui.GenericObjectEditor

weka.classifiers.lazy.IBk

About

K-nearest neighbours classifier. [More](#)

KNN 3

crossValidate False

debug False

distanceWeighting No distance weighting

meanSquared False

noNormalization False

windowSize 0

Open... Save... OK Cancel

Weka → classifiers → rules → ZeroR

