

Numeric prediction in Weka

2.12.2010

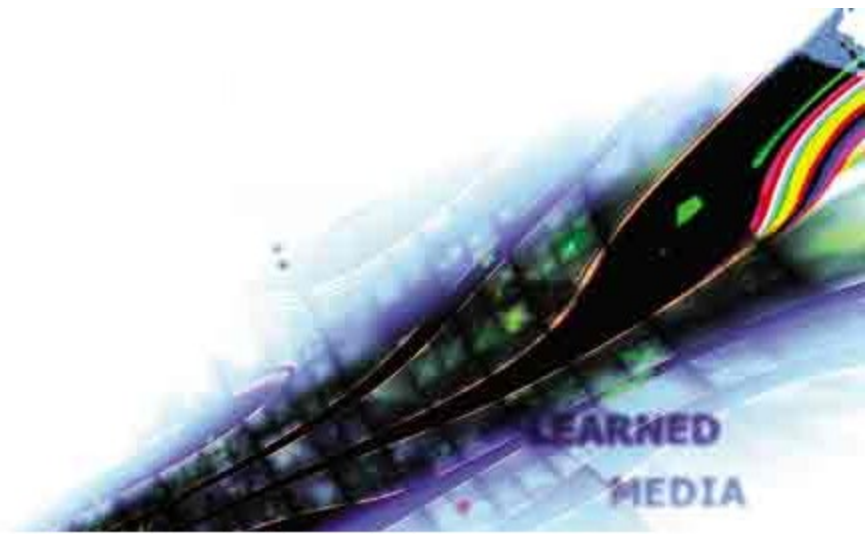
Petra Kralj Novak

Petra.Kralj.Novak@ijs.si



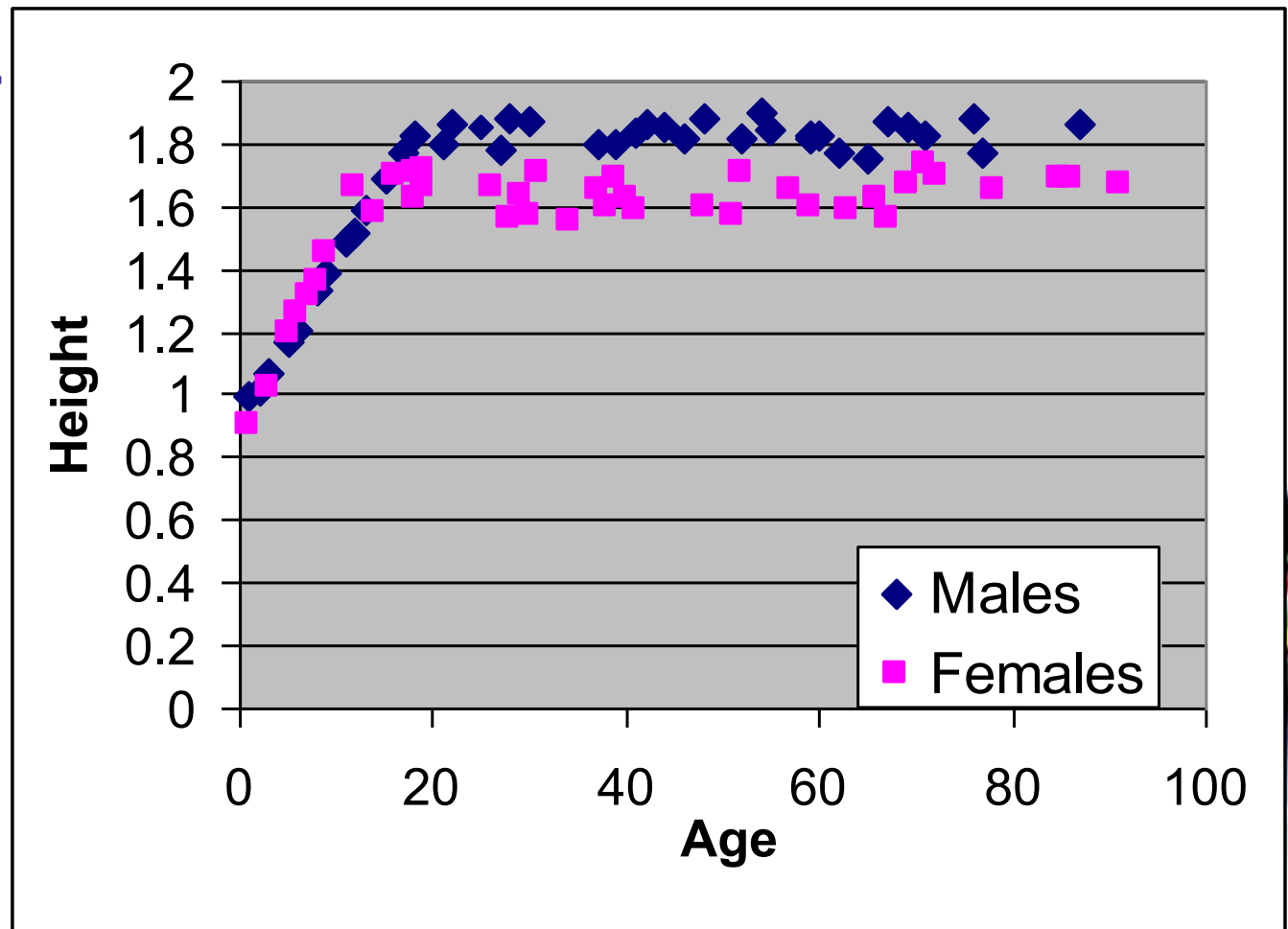
Numeric prediction models

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



regressionAgeHeight.csv

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



Filename:

Weka Explorer

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

Open file... | Open URL... | Open DB... | Undo | Edit... | Save...

Filter: Choose **None** Apply

Current relation: Relation: regressionAheHeight, Instances: 80, Attributes: 3

Attributes: All | None | Invert

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

Remove

Selected attribute: Name: Height, Missing: 0 (0%), Distinct: 80, Type: Numeric, Unique: 80 (100%)

Statistic	Value
Minimum	0.902
Maximum	1.895
Mean	1.628
StdDev	0.236

Class: Height (Num) Visualize All

Bin Range	Frequency
0.9 - 1.0	5
1.0 - 1.1	4
1.1 - 1.2	6
1.2 - 1.3	30
1.3 - 1.4	35

Status: OK Log x 0

Visualization in Weka



Weka Explorer

Preprocess

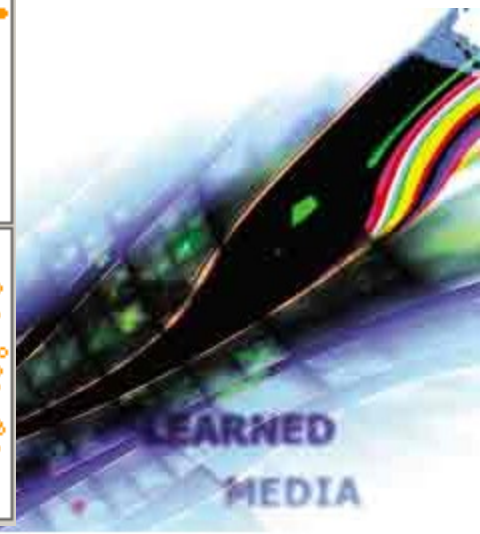
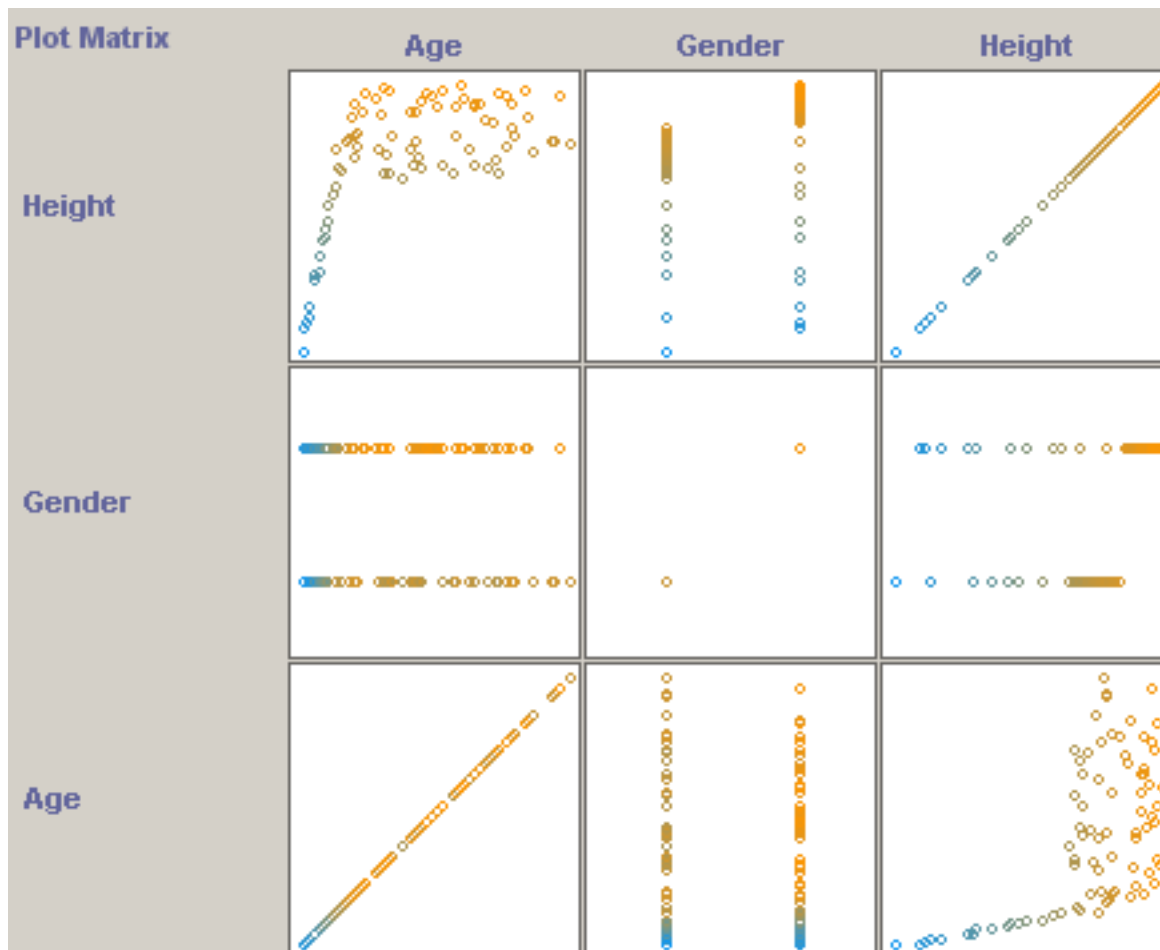
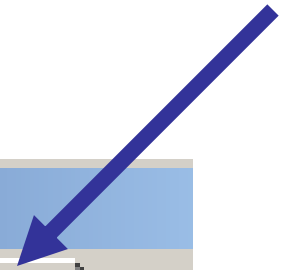
Classify

Cluster

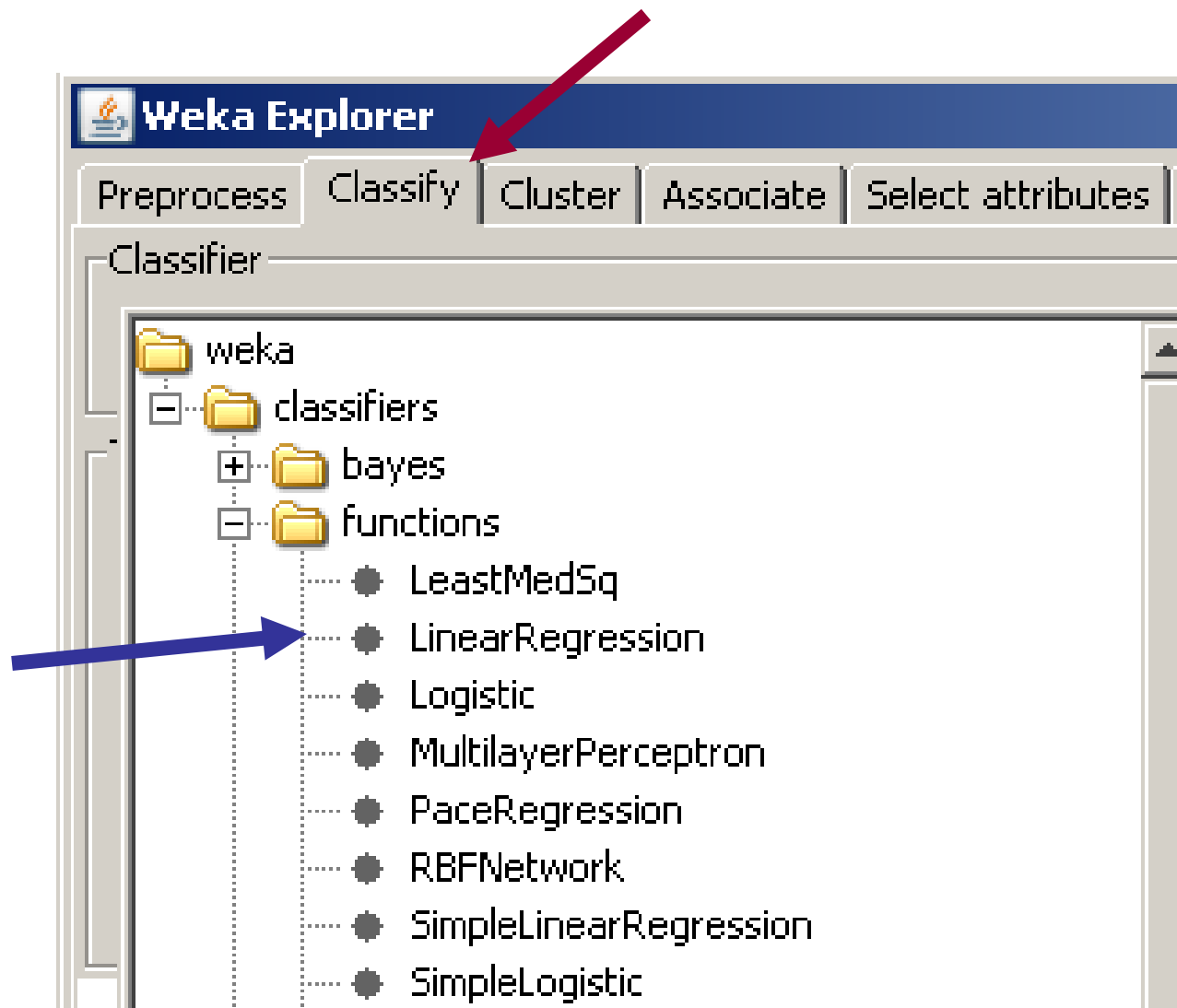
Associate

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

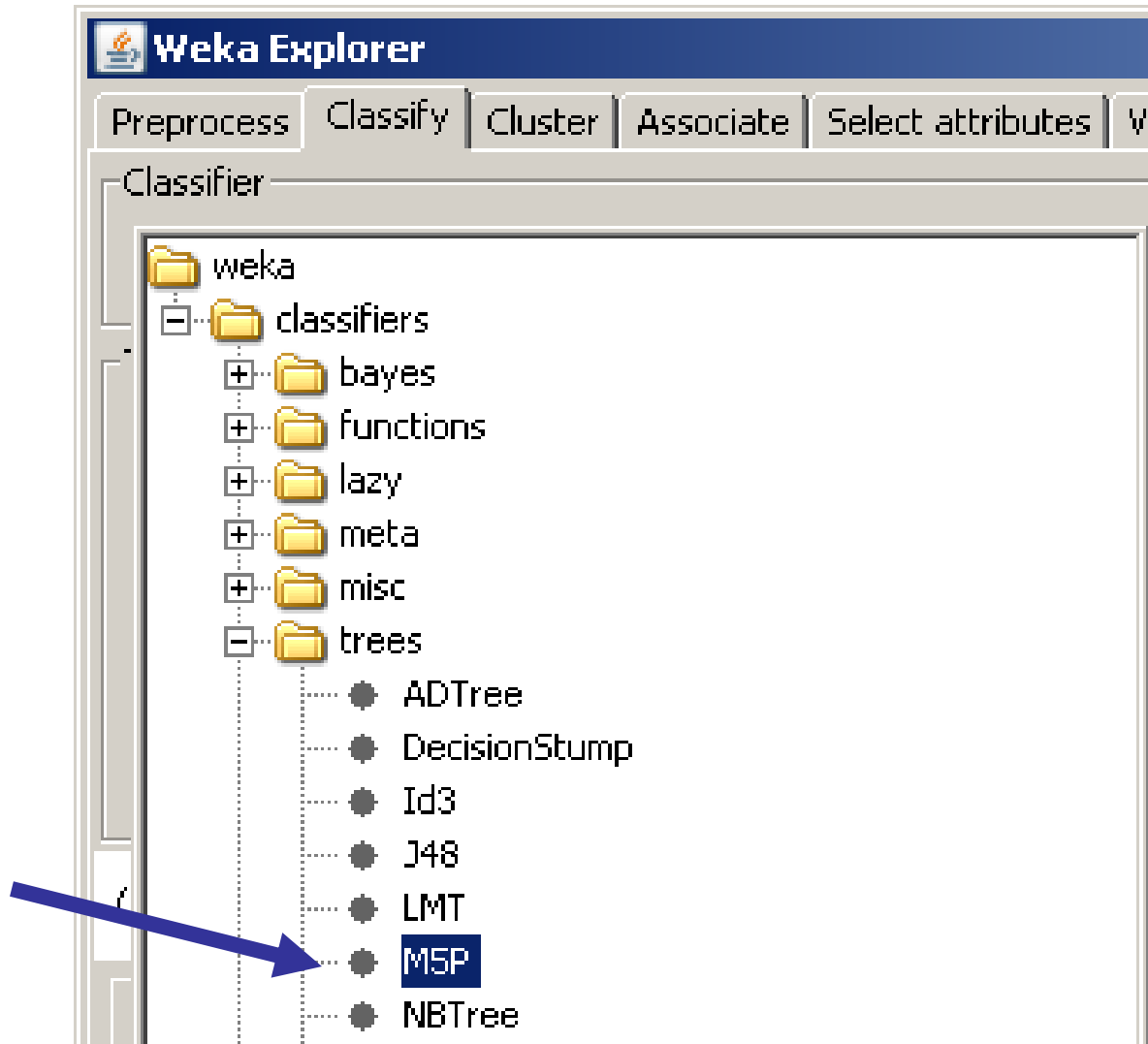
=== Cross-validation ===
=== Summary ===

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

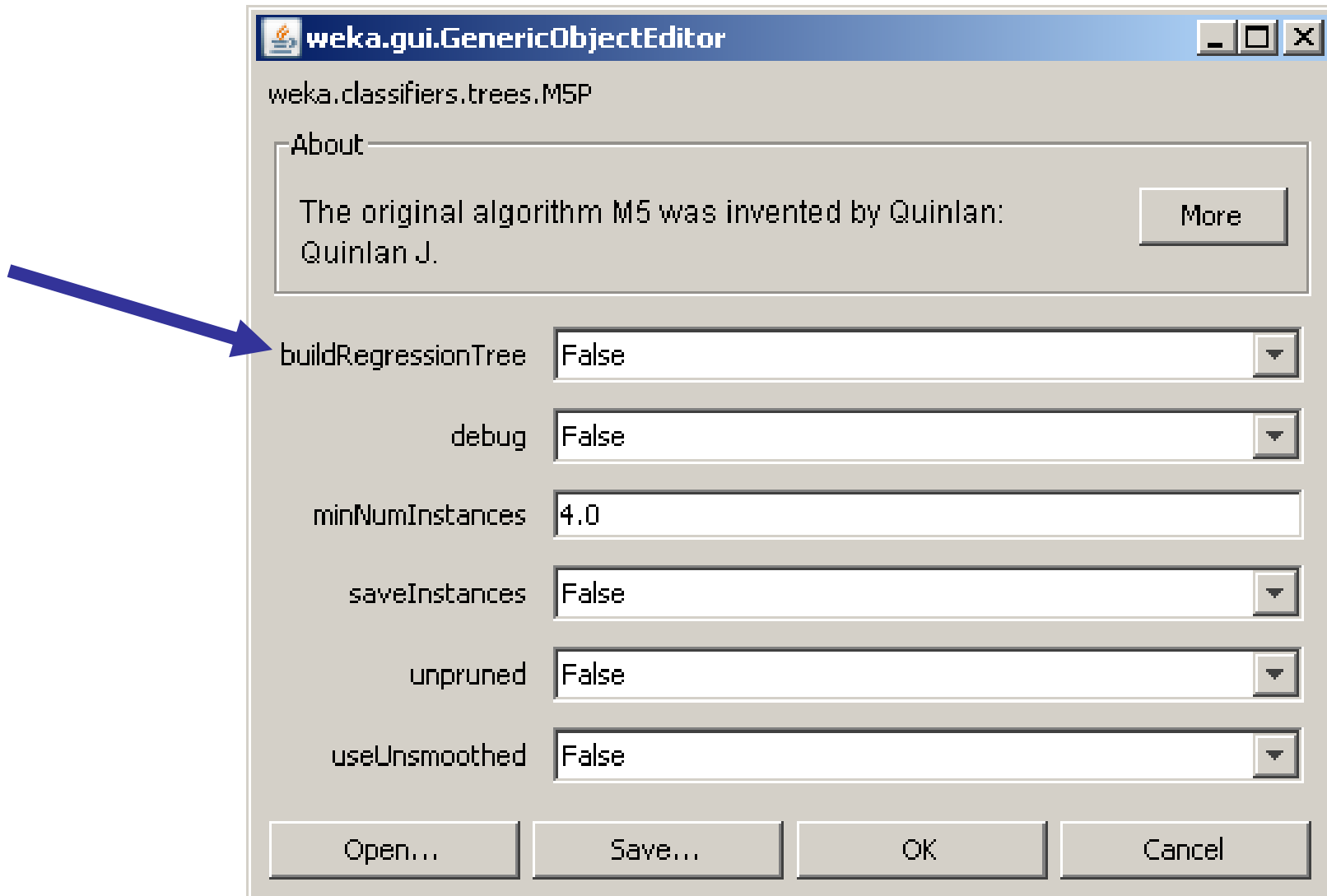
```

Status: OK  x 0

Weka → classifiers → trees → M5P



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



kNN: weka → classifiers → lazy → IBk

The image shows two windows from the Weka software. The left window, titled '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, titled 'weka.gui.GenericObjectEditor', shows the configuration for the selected classifier. A blue arrow points to the 'KNN' field, 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

