

Roadmap to M5P Classifier Output

Blackwell Data Set Example

(in_store attribute)

M5P (P stands for prime) generates M5 model trees using the M5' (read M5 prime) algorithm, introduced in 1997 by Wang & Witten. This was an enhancement of the M5 algorithm introduced in 1992 by Quinlan.

In Weka, M5P decision trees and M5 Rule-based classifiers both use M5Base as the base class (algorithmic foundation).

The M5P decision tree is also known as a "model" tree. Model trees are decision trees with linear regression functions at the leaves. They can be used to make predictions for a numeric target (class) attribute.

Intuitively this can be visualized as making a tree of linear models, where the tree structure allows linear models at the leaves to take on distinct parameter values to fit the data more closely.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MSP -M 4.0

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Num) amount

Start Stop

Result list (right-click for options)

09:25:40 - trees.MSP

Classifier output

```
=== Run information ===

Scheme:weka.classifiers.trees.MSP -M 4.0
Relation: Extended_Demographic_Data-weka.filters.unsupervised.attribute.Discretize-B5-M-1.0-R2-weka.filters.unsupervised.attribute.NumericToNominal-R1-weka.filters.unsupervised.attribute.RemoveUnnecessaryAttributes
Instances: 80000
Attributes: 5
    instore
    age
    items
    amount
    region

Test mode:10-fold cross-validation

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

M5 pruned model tree:
(using smoothed linear models)

region=1,3,4 <= 0.5 : LM1 (20000/19.779%)
region=1,3,4 > 0.5 :
| instore=0 <= 0.5 :
| | region=4 <= 0.5 :
| | | region=3,4 <= 0.5 : LM2 (16000/66.648%)
| | | region=3,4 > 0.5 : LM3 (10999/37.952%)
| | | region=4 > 0.5 : LM4 (13001/77.69%)
| | instore=0 > 0.5 : LM5 (20000/115.636%)

LM num: 1
amount =
0.537 * instore=0
- 0.0226 * age='(58.2-71.6]','(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'+
0.0567 * age='(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'
```

Colon separates tree branches (nodes) from leaf values

Leaf value is a linear model. Numbers in parenthesis indicate how much of the data is covered by this linear model

Tree branch is a splitting rule on one of the features

Status OK

Log

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **M5P -M 4.0**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Num) amount

Result list (right-click for options)

09:25:40 - trees.M5P

Classifier output

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

M5 pruned model tree:
(using smoothed linear models)

region=1,3,4 <= 0.5 : **LM1 (20000/19.779%)**

region=1,3,4 > 0.5 :

| instore=0 <= 0.5 :

| | region=4 <= 0.5 :

| | | region=3,4 <= 0.5 : LM2 (16000/66.648%)

| | | region=3,4 > 0.5 : LM3 (10999/37.952%)

| | region=4 > 0.5 : LM4 (1111/77.69%)

| instore=0 > 0.5 : LM5 (20000/115.636%)

LM num: 1

amount =

0.537 * instore=0

- 0.0226 * age='(58.2-71.6]','(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'

+ 0.0567 * age='(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'

+ 0.0149 * age='(31.4-44.8]','(-inf-31.4]'

+ 0.0217 * age='(-inf-31.4]'

+ 0.8869 * region=1,3,4

- 0.075 * region=3,4

+ 0.2005 * region=4

+ 251.5536

LM num: 2

amount =

0.4471 * instore=0

- 0.0075 * age='(58.2-71.6]','(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'

+ 227.6646 * age='(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'

+ 31.7356 * age='(31.4-44.8]','(-inf-31.4]'

Status

OK

x 0

A detailed description of the linear Model in the tree is given below. This Shows the details of the linear model Including the feature coefficients

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MSP -M 4.0

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Num) amount

Start Stop

Result list (right-click for options)

09:25:40 - trees.MSP

Classifier output

```
+ 0.2958 * region=1,3,4
- 0.2962 * region=3,4
+ 0.6893 * region=4
+ 1023.4214

LM num: 5
amount =
0.7149 * instore=0
- 0.0075 * age='(58.2-71.6]','(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'
+ 0.1141 * age='(44.8-58.2]','(31.4-44.8]','(-inf-31.4]'
+ 0.0238 * age='(31.4-44.8]','(-inf-31.4]'
+ 0.0258 * age='(-inf-31.4]'
+ 0.2958 * region=1,3,4
- 0.0974 * region=3,4
+ 0.2599 * region=4
+ 1541.8311
```

Number of Rules : 5

Time taken to build model: 424.52 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.6701
Mean absolute error	399.9901
Root mean squared error	535.3792
Relative absolute error	68.8728 %
Root relative squared error	74.227 %
Total Number of Instances	80000

Summary statistics about the model
Are displayed at the bottom of the
Model output buffer

Status
OK

Log

Explore the learned tree further:
right-click > Visualize tree
View a classifier error plot:
right-click > Visualize classifier errors

The screenshot shows the Weka Explorer interface. The 'Classifier output' window is open, displaying the results of a cross-validation process. The 'Summary' section is highlighted with a blue box, and an arrow points from it to a larger, detailed view of the summary statistics on the right.

Classifier output

```

+ 0.2995 * region=1,3,4
- 0.2962 * region=3,4
+ 0.6593 * region=4
+ 1023.4214

IN num: 5
amount =
0.7149 * instore=0
- 0.0075 * age='(58.2-71.6)', '(44.8-58.2)', '(31.4-44.8)', '(-inf-31.4)'
+ 0.1141 * age='(44.8-58.2)', '(31.4-44.8)', '(-inf-31.4)'
+ 0.0238 * age='(31.4-44.8)', '(-inf-31.4)'
+ 0.0258 * age='(-inf-31.4)'
+ 0.2650 * region=1,3,4
- 0.0974 * region=3,4
+ 0.2599 * region=4

Number of Rules : 5
Time taken to build model: 424.52 seconds

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

Correlation coefficient          0.6701
Mean absolute error            399.9901
Root mean squared error        535.3792
Relative absolute error         68.8728 %
Root relative squared error     74.227 %
Total Number of Instances      80000
  
```

Summary Statistics

== Cross-validation ==	
== Summary ==	
Correlation coefficient	0.6701
Mean absolute error	399.9901
Root mean squared error	535.3792
Relative absolute error	68.8728 %
Root relative squared error	74.227 %
Total Number of Instances	80000

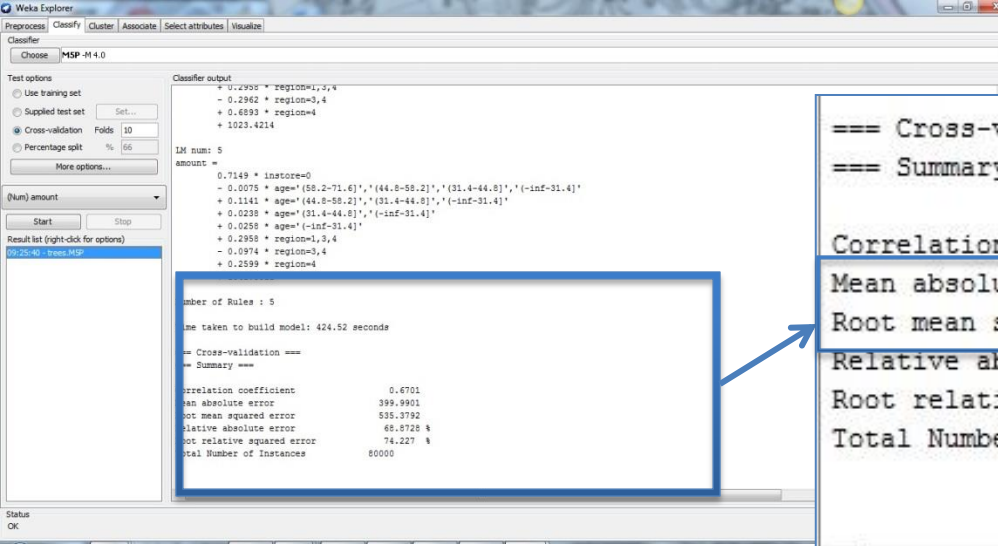
Summary Statistics

These statistics can be used to evaluate the model's success in predicting the numeric outcome.

Correlation Coefficient

In WEKA, the correlation coefficient (R, or multiple correlation coefficient) expresses the correlation between the predicted and actual class values. It is a measure of how well the predicted values from a forecast model "fit" new instances. In this example, cross validation was used.

A simple way to think about the correlation coefficient is as a score of how well the classifier's predicted labels correlate with the true labels. A score close to one indicates that the classifier predictions correlate very highly with the true labels. A score near zero means that the classifier's predictions have very little correlation with the true labels (in other words they carry no predictive value). And a score close to negative one means the classifier's predictions are consistently the opposite of the true labels. This last case still has predictive value, because if you know that the classifier consistently chooses the opposite of the true label, you can invert the classifier prediction and still get good performance.



```

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

```

Correlation coefficient	0.6701
Mean absolute error	399.9901
Root mean squared error	535.3792
Relative absolute error	68.8728 %
Root relative squared error	74.227 %
Total Number of Instances	80000

Summary Statistics (Continued)

The following statistics are different measures of error. It is useful to try to gain some insight as to what acceptable ranges of error are for the problem being addressed - particularly in the engineering domain. If that is not known, you simply want to achieve the lowest error possible and record that information for comparison with future efforts to improve it.

Mean absolute error (MAE)

The mean absolute error measures the average magnitude of errors in the set of predictions, without considering their direction. As the name suggests, MAE is the average over the all the verified predictions of the absolute values of the differences between the predicted and verified observations. All the individual differences are weighted equally in the average.

Rule of Thumb: Since this statistic measures the magnitude of error, lower values are better than a higher values. The threshold of acceptable MAE is a function of the classification task and data set being analyzed.

Root Mean-Squared Error (RMSE)

RMSE is an alternative to MAE. It also measures the average magnitude of errors in a set of predictions. Expressing this formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable. Conversely, mean-squared error tends to exaggerate the effect of outliers.

Rule of Thumb: Since this statistic measures the magnitude of error, lower values are better than higher values.

The screenshot shows the Weka Explorer window. The 'Classifier output' tab is selected, displaying the output of a cross-validation process. The output includes a list of rules, the number of rules (5), the time taken to build the model (424.52 seconds), and a summary of the cross-validation results. A blue box highlights the summary statistics, and a blue arrow points from this box to a larger, detailed summary table on the right.

Classifier output:

```

+ 0.2995 * region=1,3,4
- 0.2962 * region=3,4
+ 0.6593 * region=4
+ 1023.4214

IN num: 5
amount =
0.7149 * instore=0
- 0.0075 * age='(58.2-71.6)', '(44.8-58.2)', '(31.4-44.8)', '(-inf-31.4)'
+ 0.1141 * age='(44.8-58.2)', '(31.4-44.8)', '(-inf-31.4)'
+ 0.0238 * age='(31.4-44.8)', '(-inf-31.4)'
+ 0.0258 * age='(-inf-31.4)'
+ 0.2853 * region=1,3,4
- 0.0974 * region=3,4
+ 0.2599 * region=4

Number of Rules : 5
Time taken to build model: 424.52 seconds

Cross-validation summary
Correlation coefficient      0.6701
Mean absolute error        399.9901
Root mean squared error    535.3792
Relative absolute error     68.8728 %
Root relative squared error 74.227 %
Total Number of Instances  80000
  
```

Cross-validation Summary:

Cross-validation Summary	
Correlation coefficient	0.6701
Mean absolute error	399.9901
Root mean squared error	535.3792
Relative absolute error	68.8728 %
Root relative squared error	74.227 %
Total Number of Instances	80000

Summary Statistics (Continued)

Root Relative Squared Error

The error is made relative to what it would have been if a simple predictor would have been used. The simple predictor is just the average of the actual values for the attribute being predicted. It is calculated by taking the total of the squared error (discussed on the previous slide) and dividing it by the total squared error of the simple (default) predictor.

Rule of Thumb: Since this statistic measures the magnitude of error, lower values are better than a higher values.

Relative Absolute Error

This is statistic is similar to the Root Relative Squared Error. The absolute error is made relative to the simple predictor as with the Root Relative Squared Error. The Relative Absolute Error is calculated by taking the Mean Absolute Error (discussed in the previous slide) and dividing it by the mean absolute error of the default predictor (the average of the actual values being predicted.)

Rule of Thumb: Since this statistic measures the magnitude of error, lower values are better than a higher values.

M5P Resources:

<http://www.opentox.org/dev/documentation/components/m5p>

<http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/M5P.html>

Original academic papers:

<http://sci2s.ugr.es/keel/pdf/algorithm/congreso/1992-Quinlan-AI.pdf>

<http://www.cs.waikato.ac.nz/pubs/wp/1996/uow-cs-wp-1996-31.pdf>

http://homepages.inf.ed.ac.uk/jcavazos/SMART07/paper_9_9.pdf

Read section: 3.4

Advanced reference:

<http://link.springer.com/article/10.1023%2FA%3A1007421302149>