



Predict with WEKA

Mirco Kocher
mirco.kocher@unine.ch

October 10th, 2016



On Problem Set 01

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds
☐ Percentage split %

(Nom) class

Classifier output

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

Patron:
Some -> T
Full -> F
None -> F

(10/12 instances correct)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 7 58.3333 %
Incorrectly Classified Instances 5 41.6667 %
Kappa statistic 0.1667
Mean absolute error 0.4167
Root mean squared error 0.6455
Relative absolute error 79.2683 %
Root relative squared error 122.7289 %
Total Number of Instances 12

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| | 0.500 | 0.333 | 0.600 | 0.500 | 0.545 | 0.169 | 0.583 | 0.550 | T |
| | 0.667 | 0.500 | 0.571 | 0.667 | 0.615 | 0.169 | 0.583 | 0.548 | F |
| Weighted Avg. | 0.583 | 0.417 | 0.586 | 0.583 | 0.580 | 0.169 | 0.583 | 0.549 | |

=== Confusion Matrix ===

a b <-- classified as

3 3 | a = T

2 4 | b = F

Result list (right-click for options)

11:05:18 - rules.OneR



WEKA test options

- Test with "Cross-validation"
 - Classifier comparison
- Test with "Use training set"
 - Data evaluation
- Test with "Supplied test set"
 - Predictive power



On Problem Set 01

Test options
☒ Use training set
☐ Supplied test set
☐ Cross-validation Folds
☐ Percentage split %

(Nom) class

Result list (right-click for options)
11:08:02 - rules.OneR

Classifier output

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

Patron:
 Some -> T
 Full -> F
 None -> F

(10/12 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances 10 83.3333 %
Incorrectly Classified Instances 2 16.6667 %
Kappa statistic 0.6667
Mean absolute error 0.1667
Root mean squared error 0.4082
Relative absolute error 33.3333 %
Root relative squared error 81.6497 %
Total Number of Instances 12

=== Detailed Accuracy By Class ===

 TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
 0.667 0.000 1.000 0.667 0.800 0.707 0.833 0.833 T
 1.000 0.333 0.750 1.000 0.857 0.707 0.833 0.750 F
Weighted Avg. 0.833 0.167 0.875 0.833 0.829 0.707 0.833 0.792

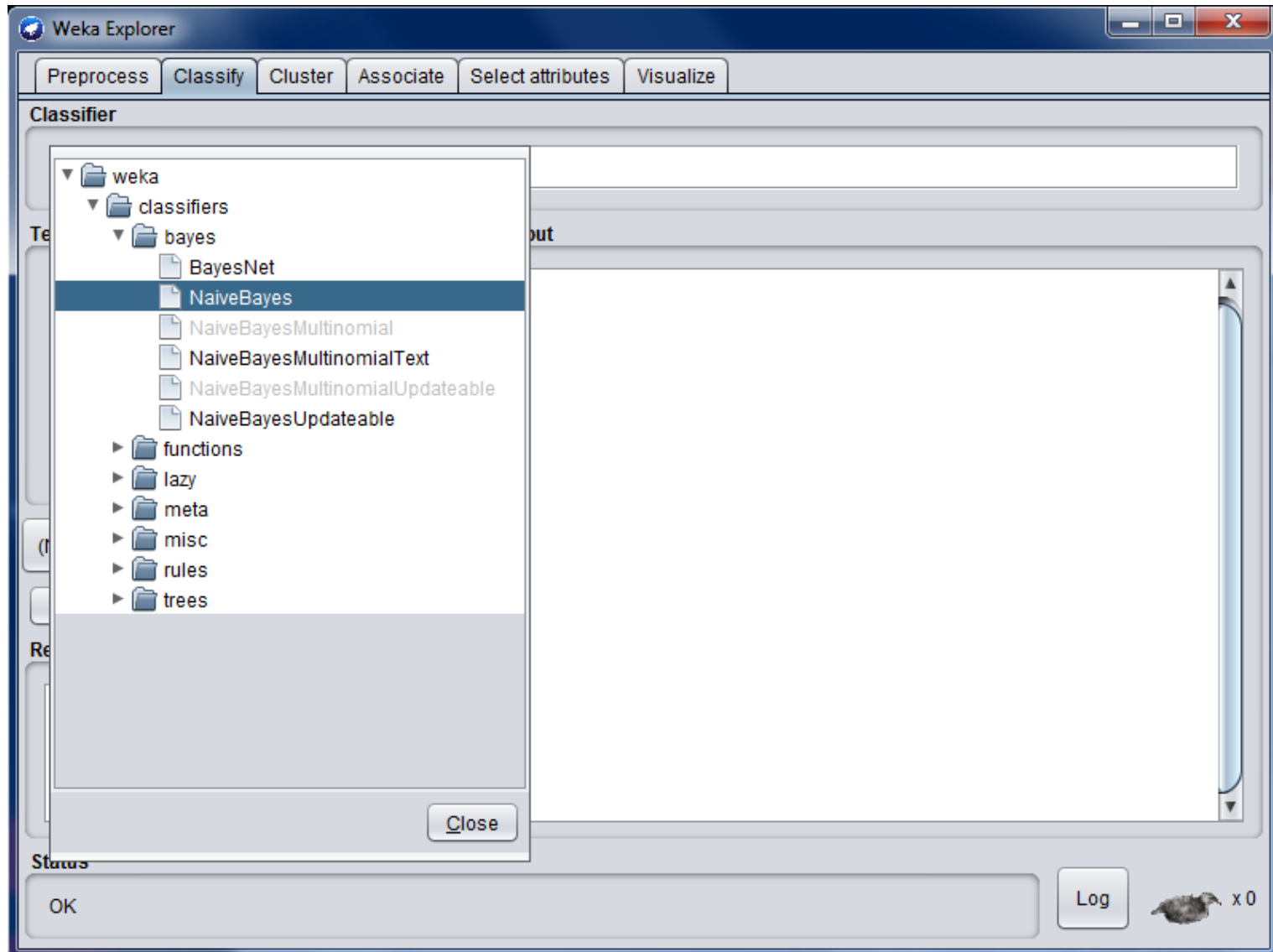
=== Confusion Matrix ===

a b <-- classified as
4 2 | a = T
0 6 | b = F





WEKA – Naïve Bayes





WEKA - Train

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☒ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:14:44 - bayes.NaiveBayes

Status

OK

Classifier output

Root relative squared error 31.7089 %

Total Number of Instances 150

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC |
|---------------|---------|---------|-----------|--------|-----------|-------|
| | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 0.960 | 0.040 | 0.923 | 0.960 | 0.941 | 0.911 |
| | 0.920 | 0.020 | 0.958 | 0.920 | 0.939 | 0.910 |
| Weighted Avg. | 0.960 | 0.020 | 0.960 | 0.960 | 0.960 | 0.940 |

=== Confusion Matrix ===

| a | b | c | <-- classified as |
|----|----|----|---------------------|
| 50 | 0 | 0 | a = Iris-setosa |
| 0 | 48 | 2 | b = Iris-versicolor |
| 0 | 4 | 46 | c = Iris-virginica |

Log x0



ARFF Sample

% Iris Plant Prediction %

@RELATION iris-predict

@ATTRIBUTE sepallength NUMERIC

@ATTRIBUTE sepalwidth NUMERIC

@ATTRIBUTE petallength NUMERIC

@ATTRIBUTE petalwidth NUMERIC

@ATTRIBUTE class {Iris-setosa, Iris-versicolor, Iris-virginica}

@DATA

4.7, 3.1, 1.2, 0.1, ?



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

1 ☒ Supplied test set **1** Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

2 More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:14:44 - bayes

- View in main window
- View in separate window
- Save result buffer
- Delete result buffer
- Load model
- Save model
- 3** Re-evaluate model on current test set
- Re-apply this model's configuration

Status

OK

Classifier evaluation options

☒ Output model

☒ Output per-class stats

☐ Output entropy evaluation measures

☒ Output confusion matrix

☒ Store predictions for visualization

☐ Error plot point size proportional to margin

Output predictions **2** Choose PlainText

☐ Cost-sensitive evaluation Set...

Random seed for XVal / % Split 1

☐ Preserve order for % Split

source code WekaClassifier

Evaluation metrics...

OK



WEKA – Predict

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☒ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:14:44 - bayes.NaiveBayes

Classifier output

```
User supplied test set
Relation:    iris-predict
Instances:   unknown (yet). Reading incrementally
Attributes:  5

=== Predictions on user test set ===

inst#    actual  predicted error prediction
   1      1:1  1:Iris-setosa      1

=== Summary ===

Total Number of Instances      0
Ignored Class Unknown Instances      1

=== Detailed Accuracy By Class ===
```

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC |
|--|---------|---------|-----------|--------|-----------|-------|
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Status

OK

Log x0



WEKA – Predict

- Train with whole training set
- Create and supply test set/sample
- Enable output prediction
- Re-evaluate model
- Get result



Problem Set 02

- Solve a decision problem by hand or WEKA with 1R
 - Predict a class using Naïve Bayes by hand
 - Predict a class using Naïve Bayes with WEKA
-
- Deadline: October 17th, 2016 at 8:00 AM



Questions?
