

CO544 – Machine Learning and Data Mining

Lab 06 – Part 2

E/17/407

WIJESOORIYA H.D

(01) attributes and values

Attribute	No of distinct records
animal	80
hair	2
feathers	2
eggs	2
milk	2
airborne	2
aquatic	2
predator	2
toothed	2

Attribute	No of distinct records
backbone	2
breathes	2
venomous	2
fins	2
legs	5
tail	2
domestic	2
catsize	2
type	7

Exercise

(01) output of the algorithm with training data set

Classifier

Choose J48 -C 0.25 -M 2

Test options

☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) type

Start Stop

Result list (right-click for options)

16:11:15 - trees.Id3
16:26:50 - rules.OneR
16:46:04 - trees.RandomForest
17:31:50 - trees.J48

Classifier output

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	80	98.7654 %
Incorrectly Classified Instances	1	1.2346 %
Kappa statistic	0.9831	
Mean absolute error	0.0062	
Root mean squared error	0.0556	
Relative absolute error	2.9101 %	
Root relative squared error	17.1647 %	
Total Number of Instances	81	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	mammal
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	bird
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	reptile
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	fish
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	amphibian
	0.833	0.000	1.000	0.833	0.909	0.907	0.992	0.910	insect
	1.000	0.014	0.875	1.000	0.933	0.929	0.993	0.875	invertebrate
Weighted Avg.	0.988	0.001	0.989	0.988	0.988	0.987	0.999	0.983	

=== Confusion Matrix ===

```
a b c d e f g <-- classified as
36 0 0 0 0 0 0 | a = mammal
0 16 0 0 0 0 0 | b = bird
0 0 3 0 0 0 0 | c = reptile
0 0 0 10 0 0 0 | d = fish
0 0 0 0 3 0 0 | e = amphibian
0 0 0 0 0 5 1 | f = insect
0 0 0 0 0 0 7 | g = invertebrate
```

Different error estimates

Kappa statistic	0.9831
Mean absolute error	0.0062
Root mean squared error	0.0556
Relative absolute error	2.9101 %
Root relative squared error	17.1647 %
Total Number of Instances	81

Percentage of classifications and misclassifications

Correctly Classified Instances	80	98.7654 %
Incorrectly Classified Instances	1	1.2346 %

Percentage of :

Classifications : 98.7654% (80 instances)

Misclassification : 1.2346% (1 instance)

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	mammal
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	bird
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	reptile
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	fish
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	amphibian
	0.833	0.000	1.000	0.833	0.909	0.907	0.992	0.910	insect
	1.000	0.014	0.875	1.000	0.933	0.929	0.993	0.875	invertebrate
Weighted Avg.	0.988	0.001	0.989	0.988	0.988	0.987	0.999	0.983	

=== Confusion Matrix ===

```
a b c d e f g <-- classified as
36 0 0 0 0 0 0 | a = mammal
 0 16 0 0 0 0 0 | b = bird
 0 0 3 0 0 0 0 | c = reptile
 0 0 0 10 0 0 0 | d = fish
 0 0 0 0 3 0 0 | e = amphibian
 0 0 0 0 0 5 1 | f = insect
 0 0 0 0 0 0 7 | g = invertebrate
```

80 instances are correctly classified and only one instance is misclassified.

One instance which is actually an insect is misclassified as an invertebrate and it is circled in red colour.

(02) output of the algorithm with test data set

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

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

(Nom) type

Result list (right-click for options)

16:11:15 - trees.Id3
16:26:50 - rules.OneR
16:46:04 - trees.RandomForest
17:31:50 - trees.J48
17:36:40 - trees.J48

Classifier output

```

=== Summary ===

Correctly Classified Instances      17          85  %
Incorrectly Classified Instances    3          15  %
Kappa statistic                    0.8187
K&B Relative Info Score            77.3631 %
K&B Information Score              43.1675 bits    2.1584 bits/instance
Class complexity | order 0         55.7985 bits    2.7899 bits/instance
Class complexity | scheme          2151.5779 bits  107.5789 bits/instance
Complexity improvement (sf)        -2095.7794 bits -104.789 bits/instance
Mean absolute error                 0.0464
Root mean squared error             0.1965
Relative absolute error             20.0843 %
Root relative squared error         55.849 %
Total Number of Instances          20

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          1.000    0.000    1.000     1.000    1.000     1.000    1.000    1.000    mammal
          1.000    0.000    1.000     1.000    1.000     1.000    1.000    1.000    bird
          0.000    0.000    ?         0.000    ?         ?        0.500    0.100    reptile
          1.000    0.000    1.000     1.000    1.000     1.000    1.000    1.000    fish
          1.000    0.053    0.500     1.000    0.667     0.688    0.974    0.500    amphibian
          0.500    0.000    1.000     0.500    0.667     0.688    0.944    0.667    insect
          1.000    0.118    0.600     1.000    0.750     0.728    0.941    0.600    invertebrate
Weighted Avg.   0.850    0.020    ?         0.850    ?         ?        0.934    0.792

=== Confusion Matrix ===

 a b c d e f g  <-- classified as
5 0 0 0 0 0 0 | a = mammal
0 4 0 0 0 0 0 | b = bird
0 0 0 0 1 0 1 | c = reptile
0 0 0 3 0 0 0 | d = fish
0 0 0 0 1 0 0 | e = amphibian
0 0 0 0 0 1 1 | f = insect
0 0 0 0 0 0 3 | g = invertebrate

```

Different error estimates

Kappa statistic	0.8187
Mean absolute error	0.0464
Root mean squared error	0.1965

Percentage of classifications and misclassifications

Correctly Classified Instances	17	85	%
Incorrectly Classified Instances	3	15	%

Percentage of :

Classifications : 85% (17 instances)

Misclassification : 15% (3 instance)

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	mammal
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	bird
	0.000	0.000	?	0.000	?	?	0.500	0.100	reptile
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	fish
	1.000	0.053	0.500	1.000	0.667	0.688	0.974	0.500	amphibian
	0.500	0.000	1.000	0.500	0.667	0.688	0.944	0.667	insect
	1.000	0.118	0.600	1.000	0.750	0.728	0.941	0.600	invertebrate
Weighted Avg.	0.850	0.020	?	0.850	?	?	0.934	0.792	

```
=== Confusion Matrix ===
```

```

a b c d e f g  <-- classified as
5 0 0 0 0 0 0 | a = mammal
0 4 0 0 0 0 0 | b = bird
0 0 0 0 1 0 1 | c = reptile
0 0 0 3 0 0 0 | d = fish
0 0 0 0 1 0 0 | e = amphibian
0 0 0 0 0 1 1 | f = insect
0 0 0 0 0 0 3 | g = invertebrate

```

17 instances are correctly classified and 3 instances are misclassified.

One instance which is actually an insect is misclassified as an invertebrate.

Two reptiles are misclassified as an amphibian and an invertebrate.

(03) Comments on the two results obtained in above (1) and (2)

Accuracy on training data set is higher than the accuracy on test data set. (accuracy on train data = 98.7654% , accuracy on test data = 85%). Only one instance is misclassified in the training data set and 3 instances are incorrectly classified in test data set.

(04) predictions

```

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Total Number of Instances          0
Ignored Class Unknown Instances    20

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?      ?      ?      ?      ?      ?      ?      ?      mammal
      ?      ?      ?      ?      ?      ?      ?      ?      bird
      ?      ?      ?      ?      ?      ?      ?      ?      reptile
      ?      ?      ?      ?      ?      ?      ?      ?      fish
      ?      ?      ?      ?      ?      ?      ?      ?      amphibian
      ?      ?      ?      ?      ?      ?      ?      ?      insect
      ?      ?      ?      ?      ?      ?      ?      ?      invertebrate
Weighted Avg.  ?      ?      ?      ?      ?      ?      ?      ?

=== Confusion Matrix ===

a b c d e f g  <-- Classified as
0 0 0 0 0 0 0 | a = mammal
0 0 0 0 0 0 0 | b = bird
0 0 0 0 0 0 0 | c = reptile
0 0 0 0 0 0 0 | d = fish
0 0 0 0 0 0 0 | e = amphibian
0 0 0 0 0 0 0 | f = insect
0 0 0 0 0 0 0 | g = invertebrate

```