

WEKA ΠΑΡΑΔΕΙΓΜΑΤΑ

Αναπλ. Καθηγ. Στελιος Ζήμερας
Τμηση Στατιστικής και Αναλογιστικών –
Χρηματοοικονομικών Μαθηματικών
Πανεπιστήμιο Αιγαίου
Σαμος

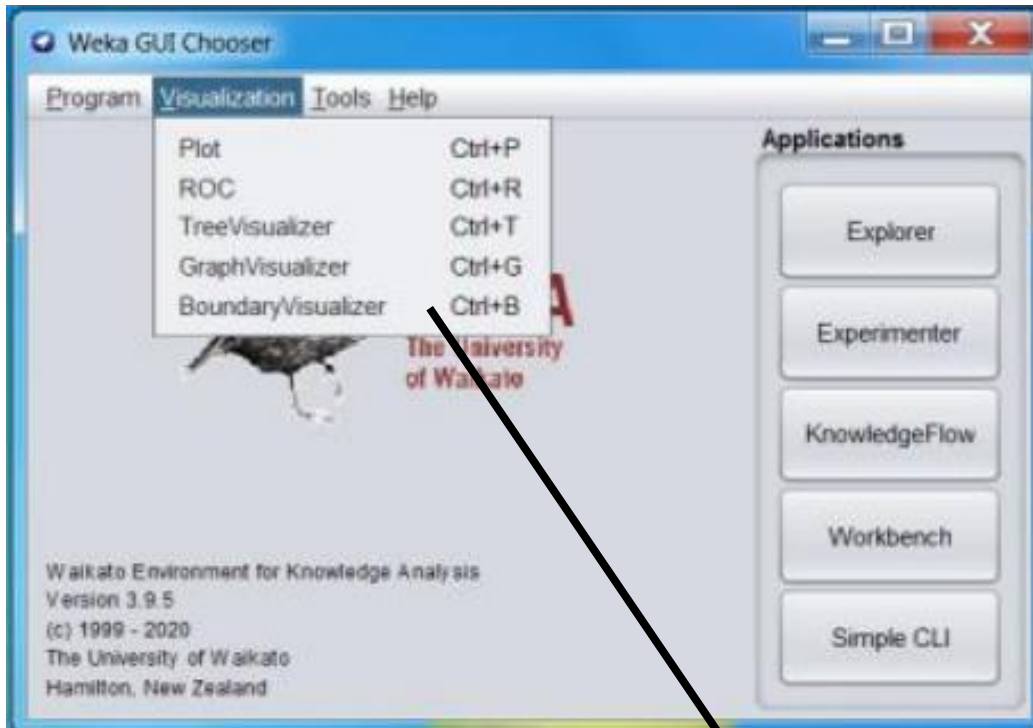
2021

Παράδειγμα



Ταχύρυθμη οπτικοποίηση δεδομένων

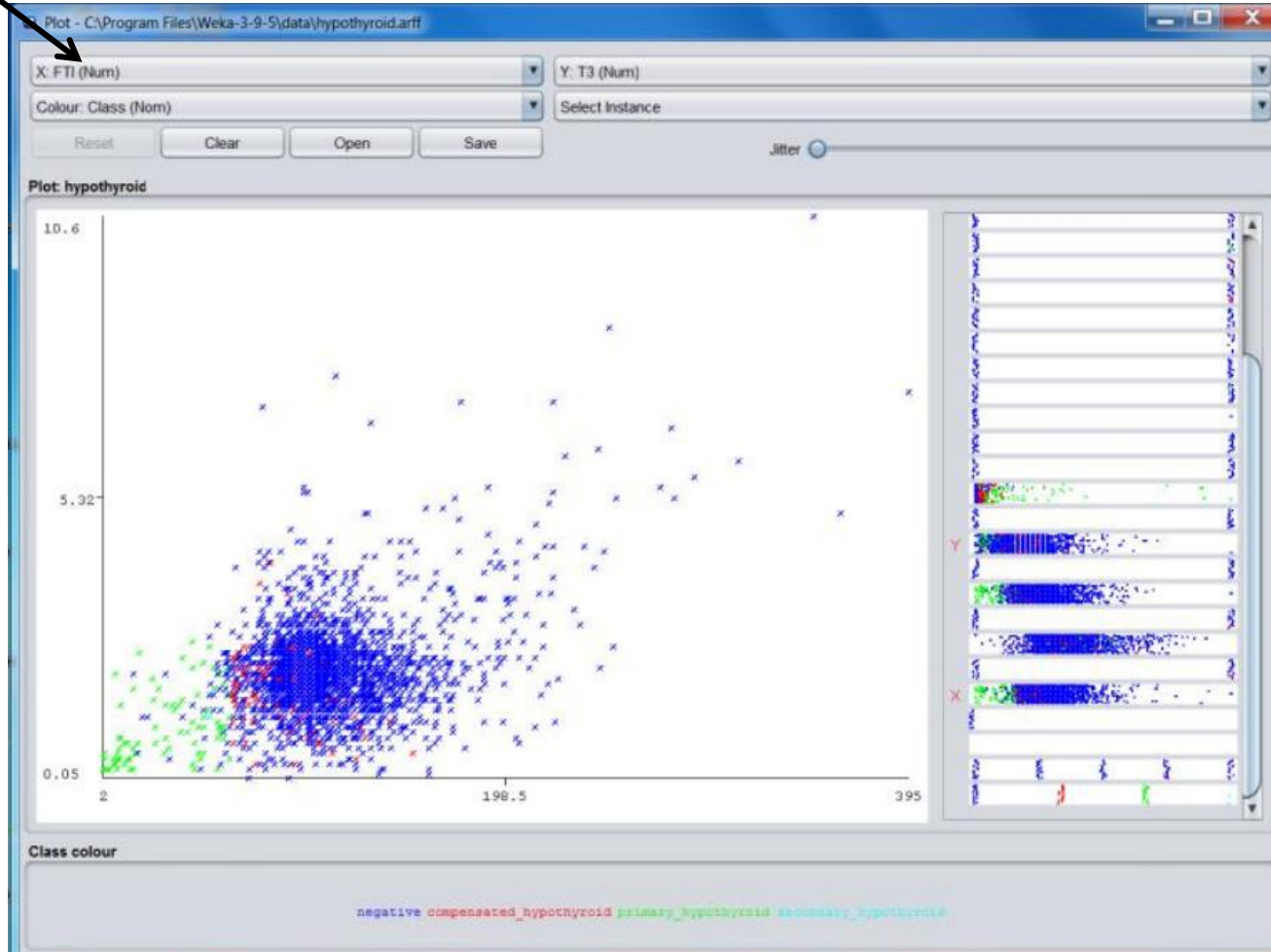
Παράδειγμα



Μέθοδοι οπτικοποίηση δεδομένων

Παράδειγμα

Plot	Ctrl+P
ROC	Ctrl+R
TreeVisualizer	Ctrl+T
GraphVisualizer	Ctrl+G
BoundaryVisualizer	Ctrl+B



Παράδειγμα

Plot	Ctrl+P
ROC	Ctrl+R
TreeVisualizer	Ctrl+T
GraphVisualizer	Ctrl+G
BoundaryVisualizer	Ctrl+B

The screenshot displays the BoundaryVisualizer application window. The interface includes several control panels:

- About:** A text field containing "Class for visualizing class probability estimates." and a "More" button.
- Dataset:** A text field with "thyroid" and an "Open File" button.
- Classifier:** A "Choose" button and a text field with "None".
- Class Attribute:** A dropdown menu showing "sex (Nom)".
- Visualization Attributes:** Two dropdown menus for "X: age (Num)" and "Y: T3 (Num)".
- Class color:** A panel with a red 'F' and a green 'M' indicator.
- Plot:** A scatter plot showing data points in red and green, forming a shape resembling a thyroid gland. The plot has axes labeled from 0 to 10.5.
- Add / remove data points:** Radio buttons for "Add points" (selected) and "Remove points", with a "Remove all" button and an "Open a new window" button.
- Sampling control:** Three input fields for "Base for sampling (r)", "Num. locations per pixel", and "Kernel bandwidth (k)", each with a value of "1".
- Plotting:** A checkbox for "Plot training data" and a "Start" button.

Weka Explorer

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

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

Filter: Choose **None** Apply Stop

Current relation
 Relation: hypothyroid | Instances: 3772 | Attributes: 30 | Sum of weights: 3772

Attributes

No.	Name
1	<input checked="" type="checkbox"/> age
2	<input type="checkbox"/> sex
3	<input type="checkbox"/> on thyroxine
4	<input type="checkbox"/> query on thyroxine
5	<input type="checkbox"/> on antithyroid medication
6	<input type="checkbox"/> sick
7	<input type="checkbox"/> pregnant
8	<input type="checkbox"/> thyroid surgery
9	<input type="checkbox"/> I131 treatment
10	<input type="checkbox"/> query hypothyroid
11	<input type="checkbox"/> query hyperthyroid
12	<input type="checkbox"/> lithium
13	<input type="checkbox"/> goitre
14	<input type="checkbox"/> tumor
15	<input type="checkbox"/> hypopituitary
16	<input type="checkbox"/> psych
17	<input type="checkbox"/> TSH measured
18	<input type="checkbox"/> ...

Selected attribute
 Name: age | Missing: 1 (0%) | Distinct: 93 | Type: Numeric | Unique: 5 (0%)

Statistic	Value
Minimum	1
Maximum	455
Mean	51.736
StdDev	20.085

Class: Class (Nom)

Status
 x 0

Weka Explorer

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

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

Filter: Choose New

Current relation: Relation: hyp Instances: 37

Attributes: All

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

All attributes

3728 I131 treatment 43	3525 query hypothyroid 147	3719 query hyperthyroid 53	3719 lithium 53
3713 goitre 58	3538 tumor 234	3535 hypopituitary 237	3754 psych 18
3738 TSH measured 34	3576 TSH 95	3771 T3 measured 1	3580 T3 184
3403 TT4 measured 268	0.01 265 530 TT4 2 216 430	3003 T4U measured 790	0.05 5.33 10.6 T4U 0.25 1.28 2.32
3541 FTI measured 231	FT 0 1	3395 TBC measured 387	TBC 0 1

OK | Log | x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -A -B -I -R -output

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)

- 20:28:18 - functions.MultilayerPerceptron
- 20:29:13 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron

Status

OK

weka.gui.GenericObjectEditor

weka.classifiers.functions.MultilayerPerceptron

About

A classifier that uses backpropagation to learn a multi-layer perceptron to classify instances. More Capabilities

GUI True

autoBuild False

batchSize 100

debug True

decay False

doNotCheckCapabilities False

hiddenLayers a

learningRate 0.3

momentum 0.2

nominalToBinaryFilter False

normalizeAttributes False

normalizeNumericClass True

numDecimalPlaces 2

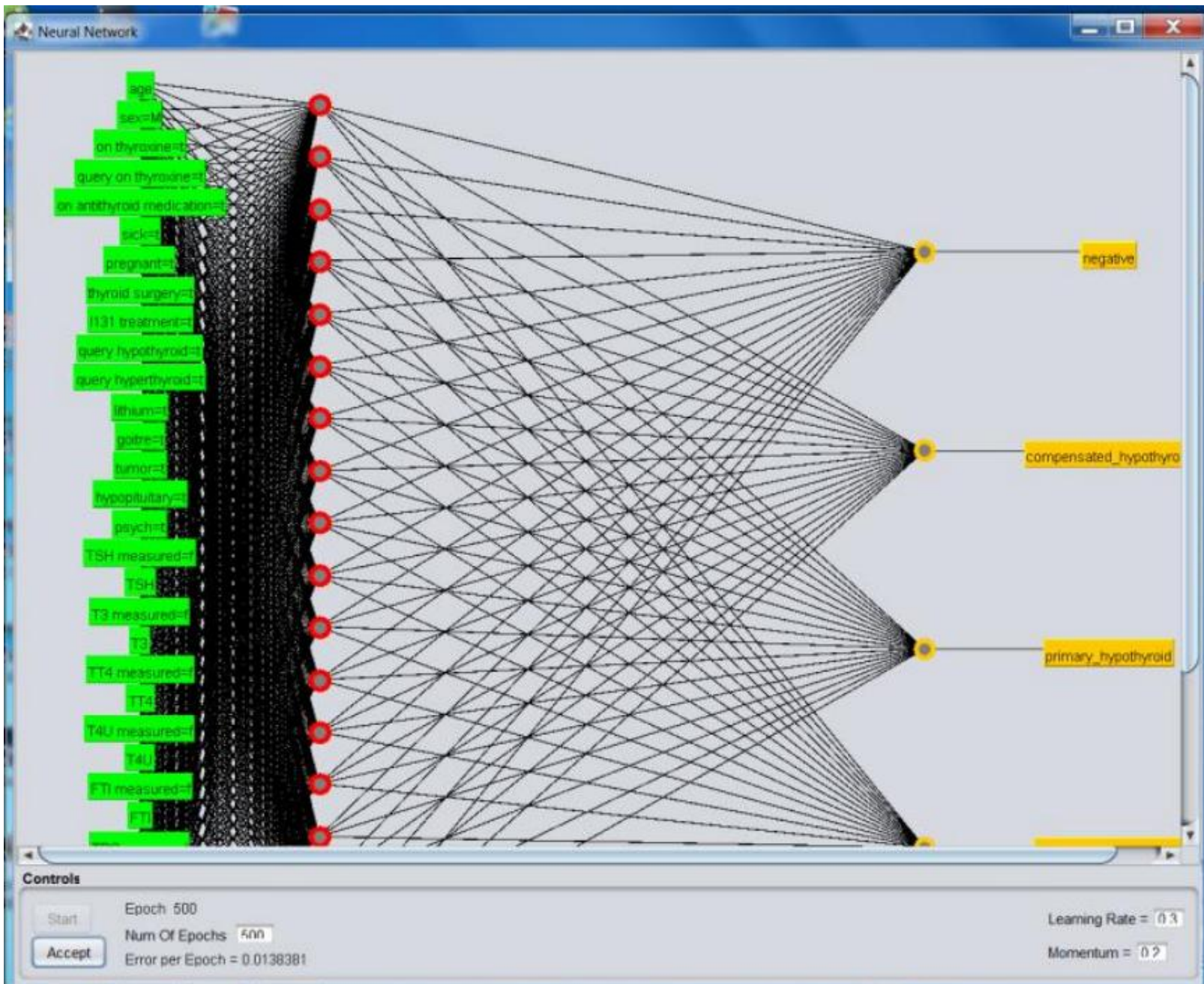
reset False

resume False

seed 0

If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime)

Open... Save... OK Cancel Log x 0



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: **MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds: 10
- Percentage split %: 00

(Nom) Class

Result list (right-click for options)

- 20:48:45 - functions.MultilayerPerceptron
- 20:49:11 - functions.MultilayerPerceptron
- 20:50:25 - functions.MultilayerPerceptron
- 20:53:20 - functions.MultilayerPerceptron**

Classifier output

```

==== Evaluation on training set ====

Time taken to test model on training data: 0.07 seconds

==== Summary ====

Correctly Classified Instances      3588          95.122 %
Incorrectly Classified Instances    184           4.878 %
Kappa statistic                    0.59
Mean absolute error                 0.0445
Root mean squared error             0.1443
Relative absolute error             61.0301 %
Root relative squared error         75.7812 %
Total Number of Instances          3772

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

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.991  0.495  0.960  0.991  0.975  0.624  0.910  0.990  negative
0.253  0.007  0.653  0.253  0.364  0.388  0.866  0.408  compensated_hypothyroid
0.937  0.004  0.864  0.937  0.899  0.897  0.998  0.948  primary_hypothyroid
0.000  0.000  ?  0.000  ?  ?  0.962  0.010  secondary_hypothyroid
Weighted Avg.  0.951  0.457  ?  0.951  ?  ?  0.910  0.959

==== Confusion Matrix ====

 a  b  c  d  <-- classified as
3450 24  7  0 |  a = negative
138  49  7  0 |  b = compensated_hypothyroid
 5  1 89  0 |  c = primary_hypothyroid
 1  1  0  0 |  d = secondary_hypothyroid

```

Status

OK

Weka Classifier Visualize: ThresholdCurve. (Class value negative)

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

Reset

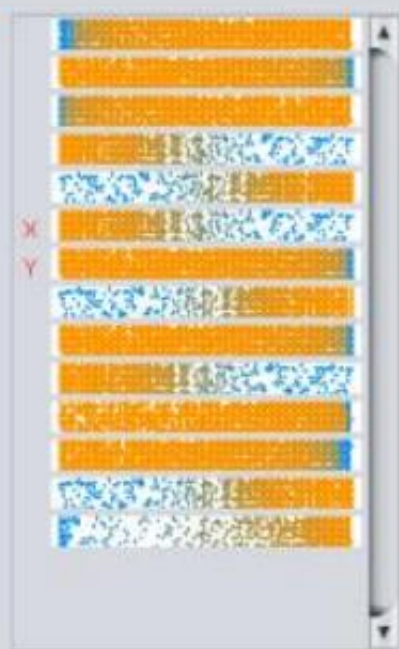
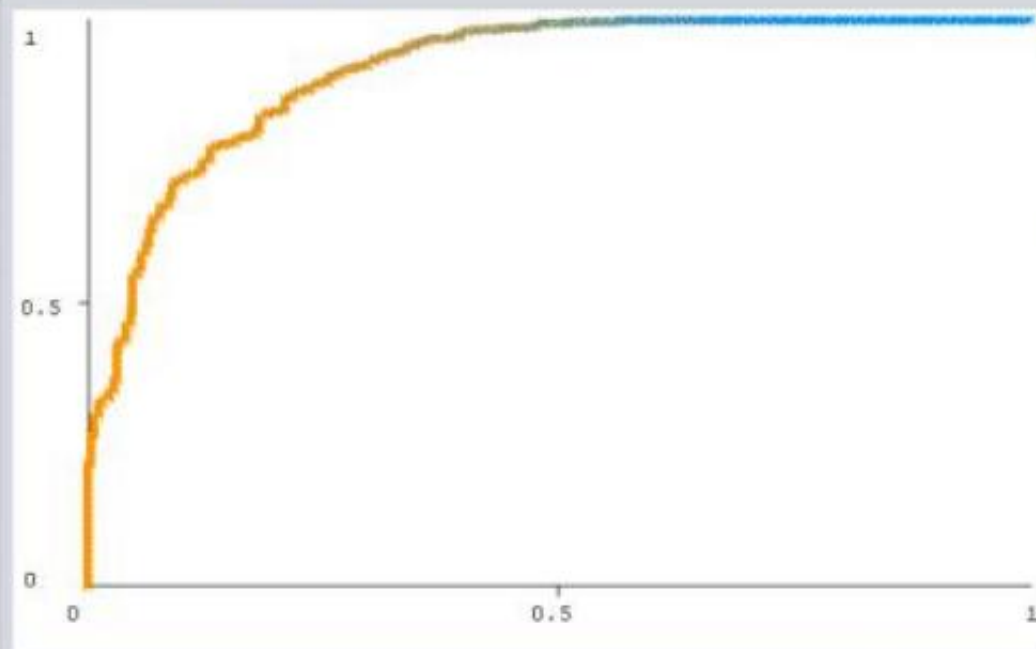
Clear

Open

Save

Jitter

Plot (Area under ROC = 0.9102)



Class colour

0.00015

0.5

1

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R first-last"

Cluster mode

- Use training set
- Supplied test set
- Percentage split %
- Classes to clusters evaluation (Nom) Class
- Store clusters for visualization

Ignore attributes

Start

Result list (right-click for options)

Clusterer output

weka.gui.GenericObjectEditor

weka.clusterers.HierarchicalClusterer

About

Hierarchical clustering class.

debug

distanceFunction

distancelsBranchLength


doNotCheckGapabilities

linkType

numClusters

printNewick

Status

OK  x 0

Clusterer
 Choose **SimpleKMeans** -init 0 -C -max-candidates 100 -periodic-pruning 100 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka core EuclideanDistance -R first-last" -I 50 -O -num-slots 1 -S 10 -output-debug-info

Cluster mode

Use training set

Supplied test set

Percentage split % 60

Classes to clusters evaluation
 (Num) Class

Store clusters for visualization

- Result list (right-click for options)**
- 21 06 33 - HierarchicalClusterer
 - 21 07 23 - SimpleKMeans
 - 21 08 14 - FarthestFirst
 - 21 08 35 - EM
 - 21 08 49 - EM
 - 21 09 16 - Cobweb
 - 21 10 05 - MakeDensityBasedClusterer
 - 21 12 07 - SimpleKMeans**

Clusterer output

```

query-hypothyroid
lithium          f          f          f
goitre           f          f          f
tumor            f          f          f
hypopituitary   f          f          f
psych            f          f          f
TSH measured    t          t          t
TSH              5.0869    1.7273    6.1273
T3 measured     t          t          t
T3              2.0135    2.5923    1.8342
TT4 measured    t          t          t
TT4             108.3193  151.9865  94.7946
T4U measured    t          t          t
T4U             0.995     1.1534    0.9459
FTI measured    t          t          t
FTI             110.4696  136.134   102.5208
TBG measured    f          f          f
TBG             0          0          0
referral source other      other      other
Class           negative  negative  negative

Time taken to build model (full training data) : 0.3 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      892 ( 24%)
1     2880 ( 76%)
    
```

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose BayesNet-D -Q weka.classifiers.bayes.net.search.local.SimulatedAnnealing -- -A 10.0 -U 10000 -D 0.999 -R 1 -S BAYES-E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Test options

Use training set
 Supplied test set
 Cross-validation Folds 10
 Percentage split % 66

(Norm) Class

Result list (right-click for options)

- 20:48:45 - functions MultilayerPerceptron
- 20:49:11 - functions MultilayerPerceptron
- 20:50:25 - functions MultilayerPerceptron
- 20:53:20 - functions MultilayerPerceptron
- 21:15:23 - bayes BayesNet
- 21:15:53 - bayes BayesNet
- 21:16:38 - bayes BayesNet
- 21:22:42 - bayes BayesNet

Classifier output

```

=== Evaluation on training set ===

Time taken to test model on training data: 0.26 seconds

=== Summary ===

Correctly Classified Instances      3754           99.5228 %
Incorrectly Classified Instances     18             0.4772 %
Kappa statistic                     0.9678
Mean absolute error                  0.0055
Root mean squared error              0.0497
Relative absolute error              7.4927 %
Root relative squared error          26.0897 %
Total Number of Instances           3772

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCG      ROC Area  PRC Area  Class
0.996  0.007  0.999  0.996  0.998  0.969  0.998  1.000  negative
1.000  0.003  0.956  1.000  0.977  0.976  0.999  0.959  compensated_hypothyroid
0.989  0.002  0.931  0.989  0.959  0.959  0.999  0.973  primary_hypothyroid
0.000  0.000  ?  0.000  ?  ?  0.998  0.148  secondary_hypothyroid
Weighted Avg.  0.995  0.007  ?  0.995  ?  ?  0.998  0.997

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
3466  8  7  0 |  a = negative
  0 194  0  0 |  b = compensated_hypothyroid
  0  1  94  0 |  c = primary_hypothyroid
  2  0  0  0 |  d = secondary_hypothyroid

```

Status

OK x 0

