Data analysis by density based clustering

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

Clusterer

Choose | MakeDensityBasedClusterer -M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode
- Use training set
- Supplied test set   Set...
- Percentage split   % 66
- Classes to clusters evaluation
  (Nom) Class
- ✓ Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)
13:02:37 - MakeDensityBasedClusterer

Clusterer output

```
=== Run information ===

Scheme:     weka.clusterers.MakeDensityBasedClusterer -M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-de:
Relation:   breast-cancer
Instances:  286
Attributes: 10
            age
            menopause
            tumor-size
            inv-nodes
            node-caps
            deg-malig
            breast
            breast-quad
            irradiat
            Class
Test mode:  evaluate on training data


=== Clustering model (full training set) ===

MakeDensityBasedClusterer:

Wrapped clusterer:
kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 1177.0

Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode
```


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

Clusterer

Choose | MakeDensityBasedClusterer -M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode
- Use training set
- Supplied test set   Set...
- Percentage split   % 66
- Classes to clusters evaluation
  (Nom) Class
- ✓ Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)
13:02:37 - MakeDensityBasedClusterer

Clusterer output

```
Number of iterations: 3
Within cluster sum of squared errors: 1177.0

Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:
                                               Cluster#
Attribute                 Full Data                   0                    1
                            (286.0)               (225.0)               (61.0)
==============================================================================
age                           50-59                 50-59                40-49
menopause                   premeno               premeno              premeno
tumor-size                    30-34                 25-29                30-34
inv-nodes                       0-2                   0-2                  0-2
node-caps                        no                    no                  yes
deg-malig                         2                     2                    3
breast                         left                  left                 left
breast-quad                left_low              left_low             left_low
irradiat                         no                    no                   no
Class           no-recurrence-events  no-recurrence-events  recurrence-events


Fitted estimators (with ML estimates of variance):

Cluster: 0 Prior probability: 0.7847

Attribute: age
Discrete Estimator. Counts =  1 2 27 65 84 46 7 1 1  (Total = 234)
Attribute: menopause
Discrete Estimator. Counts =  8 106 114  (Total = 228)
Attribute: tumor-size
```