732A75: Association Analysis-2

Liu ID: vinbe289, muhha866

The clusters in the data are found out using the different attributes by the clustering algorithms. Each instance is assigned a class here. The class assigned to each instance may either depend on the attributes in such a way that the clustering algorithm cannot relate the attributes to find out clusters corresponding to different classes or the classes assigned to each instance may not depend on the attributes at all. Besides as the name itself indicates the objective function of the clustering algorithms is not classification but lustering.

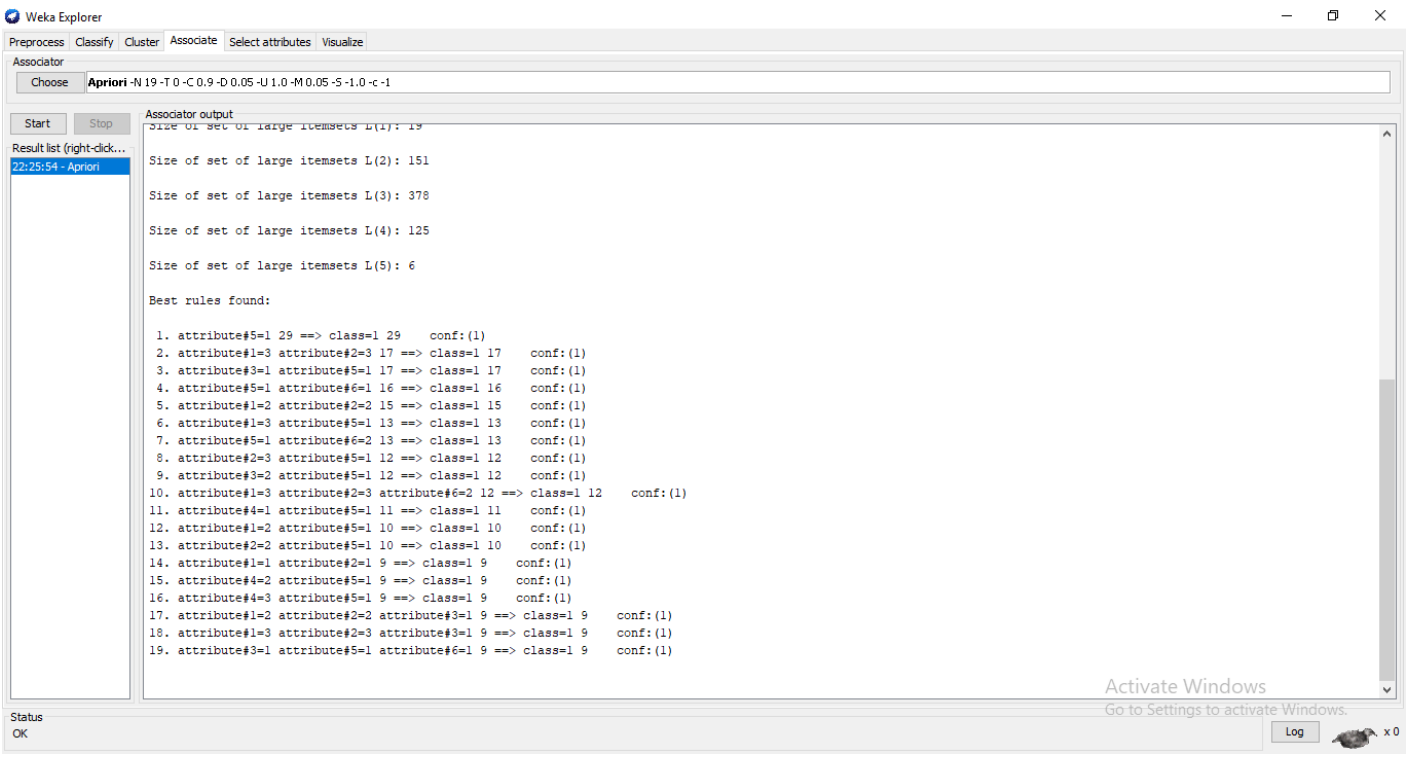
We can describe class 1 using these rules.

1. attribute#5=1 29 ==> class=1 29 conf:(1)

2. attribute#1=3 attribute#2=3 17 ==> class=1 17 conf:(1)

3. attribute#3=1 attribute#5=1 17 ==> class=1 17 conf:(1)

4. attribute#5=1 attribute#6=1 16 ==> class=1 16 conf:(1)



Consider the following generated association rules below:

attribute#1=3 attribute#2=3 17 ==> class=1 17 conf:(1)

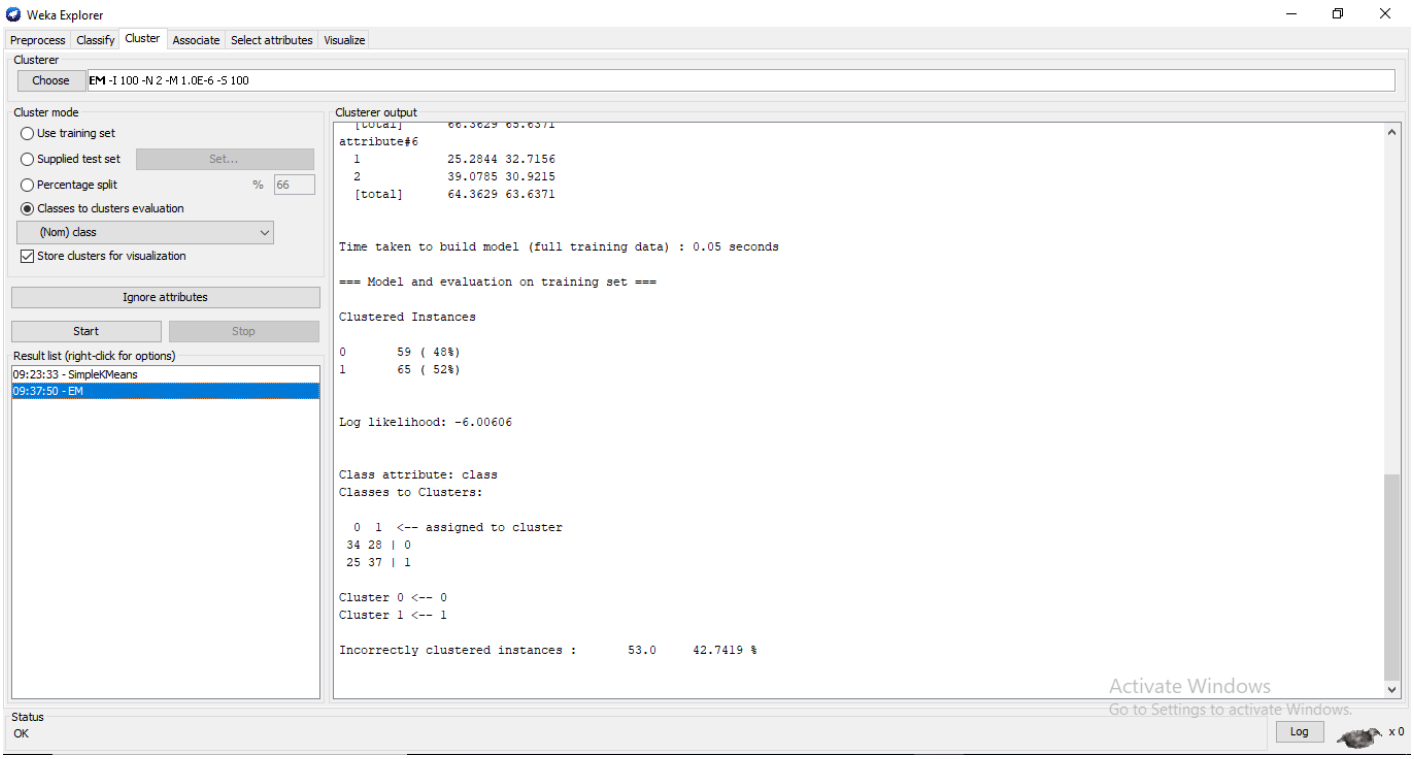
attribute#1=2 attribute#2=2 15 ==> class=1 15 conf:(1)

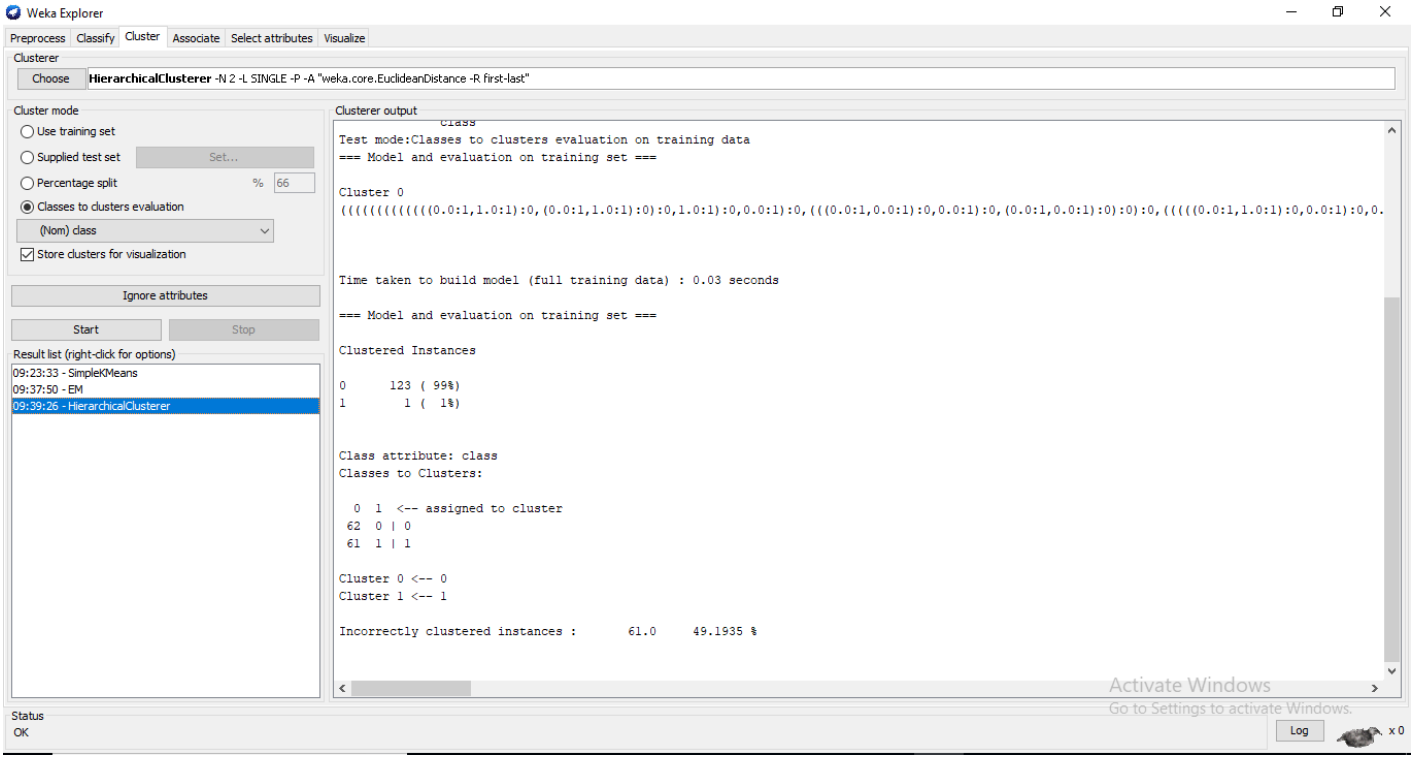
attribute#1=1 attribute#2=1 9 ==> class=1 9 conf:(1)

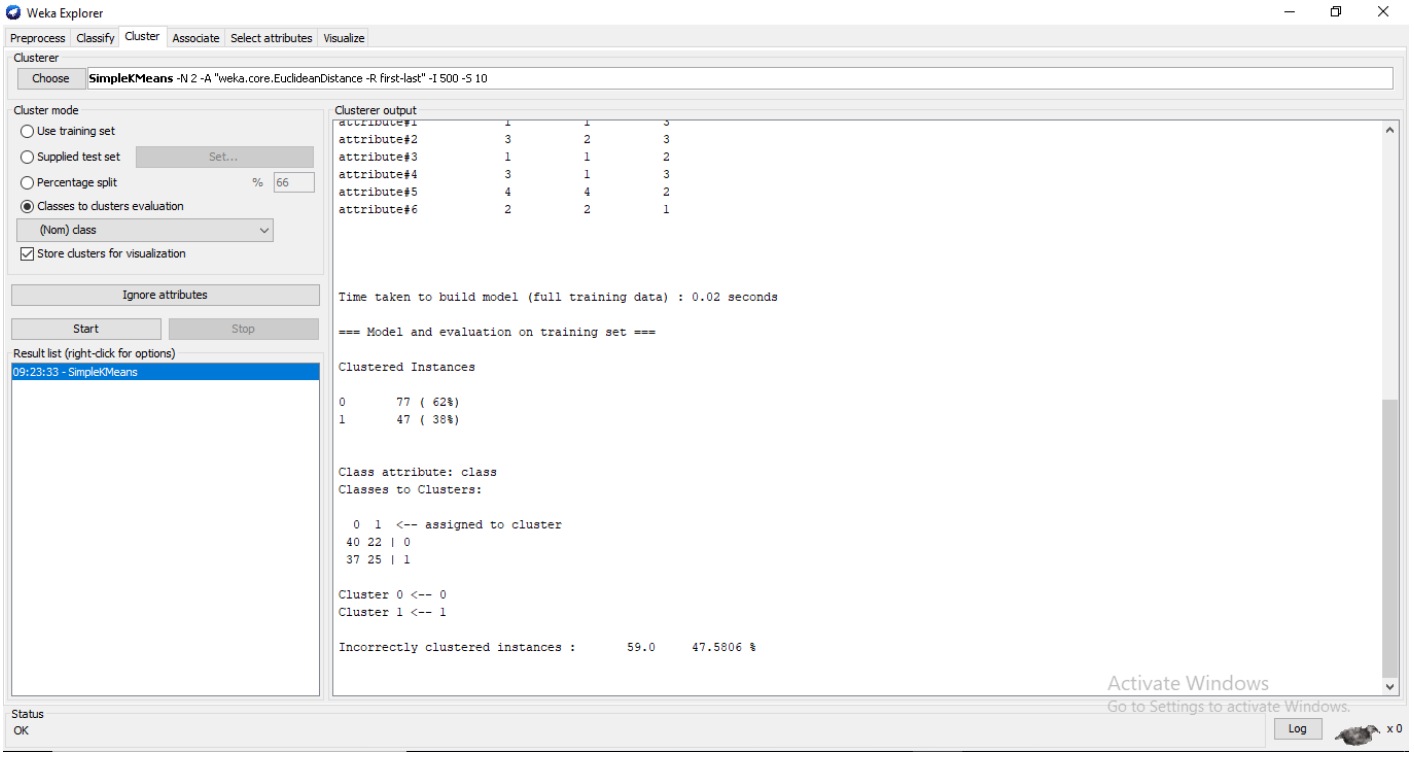
Attribute 1 and 2 has only 3 distinct values i.e. 1,2 and 3. According to the above 3 association rules generated it is evident that even when the attributes 1 and 2 takes all the 3 possible distinct values the instance is classified into class 1 which means that there is no particular rule as such that the clustering algorithms recognize to classify particular instances to class 1.

We can conclude that the clustering algorithms perform poorly on this dataset only because of the dataset itself. This statement is supported by the screenshots included for EM, SimpleKmeans and Hierarchical clustering algorithms. The number of instances wrongly classified is the sum of diagonal.

The dataset addresses classification problem but we are using it for clustering. Clustering and classification is not the same and cannot be used interchangeably.




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 Set...
Percentage split % 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start Stop
Result list (right-click for options)
09:23:33 - SimpleKMeans
09:37:50 - EM
09:39:26 - HierarchicalClusterer
Clusterer output
Test mode:Classes to clusters evaluation on training data
=== Model and evaluation on training set ===
Cluster 0
Time taken to build model (full training data) : 0.03 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 123 (99%)
1 1 (1%)
Class attribute: class
Classes to Clusters:
0 1 <-- assigned to cluster
62 0 | 0
61 1 | 1
Cluster 0 <-- 0
Cluster 1 <-- 1
Incorrectly clustered instances : 61.0 49.1935 %
Activate Windows
Go to Settings to activate Windows.
Status
OK
Log
x 0


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Clusterer
Choose SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -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)
09:23:33 - SimpleKMeans
Clusterer output
attribute#2 3 2 3
attribute#3 1 1 2
attribute#4 3 1 3
attribute#5 4 4 2
attribute#6 2 2 1
Time taken to build model (full training data) : 0.02 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 77 (62%)
1 47 (38%)
Class attribute: class
Classes to Clusters:
0 1 <-- assigned to cluster
40 22 | 0
37 25 | 1
Cluster 0 <-- 0
Cluster 1 <-- 1
Incorrectly clustered instances : 59.0 47.5806 %
Activate Windows
Go to Settings to activate Windows.
Status
OK
Log
x 0