Date :

Practical No. 6

Aim: To demonstrate hierarchical clustering on given data set.

Theory:

Clustering is one of the most common exploratory data analysis techniques used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data point in the same subgroup (cluster) are very similar while data point in different clusters are very different. In otherwords, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible acc to the similarity measure such as euclidean-based distance or correlation based distance. The decision of which similarity measure to used is application-specific.

Hierarchical clustering :→

It is also known as Hierarchical cluster analysis or HCA, is an unsupervised clustering approach that includes forming group with a top-to-bottom order.

The hierarchical clustering techniques has two approaches:

① Agglomerative : It is bottom up approach, in which the algo starts with taking all data points as single clusters and merging them until one cluster is left.

② Divisive : It is the reverse of agglomerative algo. as it is a top-down approach.

Agglomerative Hierarchical clustering :→

Agglomerative clustering or bottom up clustering, essentially started from an individual cluster (each data point is considered as an individual cluster, also called leaf), then every cluster calculates their distance with each other

 The

Date :

two clusters with the shortest distance with each other would merge creating what we called node. Newly formed clusters once again calculating the member of their cluster distance with another cluster outside of their cluster. The process is repeated until all the data points assigned to one cluster called root. The result is a tree-based representation of the objects called dendrogram.

Conclusion :

Thus we have implemented hierarchical algorithm for clustering Succesfully.

Viva Voce :

① what is hierarchical clustering Algorithm ?

→ Hierarchical clustering, also known as hierarchical cluster analysis or HCA, is an unsupervised clustering approach that include forming groups with a top-to-down (bottom) order.

② what are the various types of hierarchical clustering?

→ There are two types of hierarchical clustering, They are:

a) Agglomerative

b) Divisive.

③ what is dendrogram in HCA ?

→ Dendrogram - is a diagram that shows the hierarchical relationship betn objects.

④ Explain the different linkage method used in HCA ?

→ ① Single linkage :→

Single linkage returns minimum distance betn two point, where each points belong to two different clusters.

Date :

② Complete linkage :→
It returns the maximum distance betn each data point.

③ Average linkage :→
It returns the average of distances betn all pairs of data points.

④ Centroid linkage :→
It returns the distance betn centroid of clusters.

Weka Explorer

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

Clusterer

Choose **HierarchicalClusterer** -N 4 -L SINGLE -P -B -A "weka.core.EudideanDistance -R first-last" -do-not-check-capabilities

Cluster mode

- (•) Use training set
- ( ) Supplied test set — Set...
- ( ) Percentage split — % 66
- ( ) Classes to clusters evaluation
  (Nom) play
- [x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

23:08:38 - HierarchicalClusterer

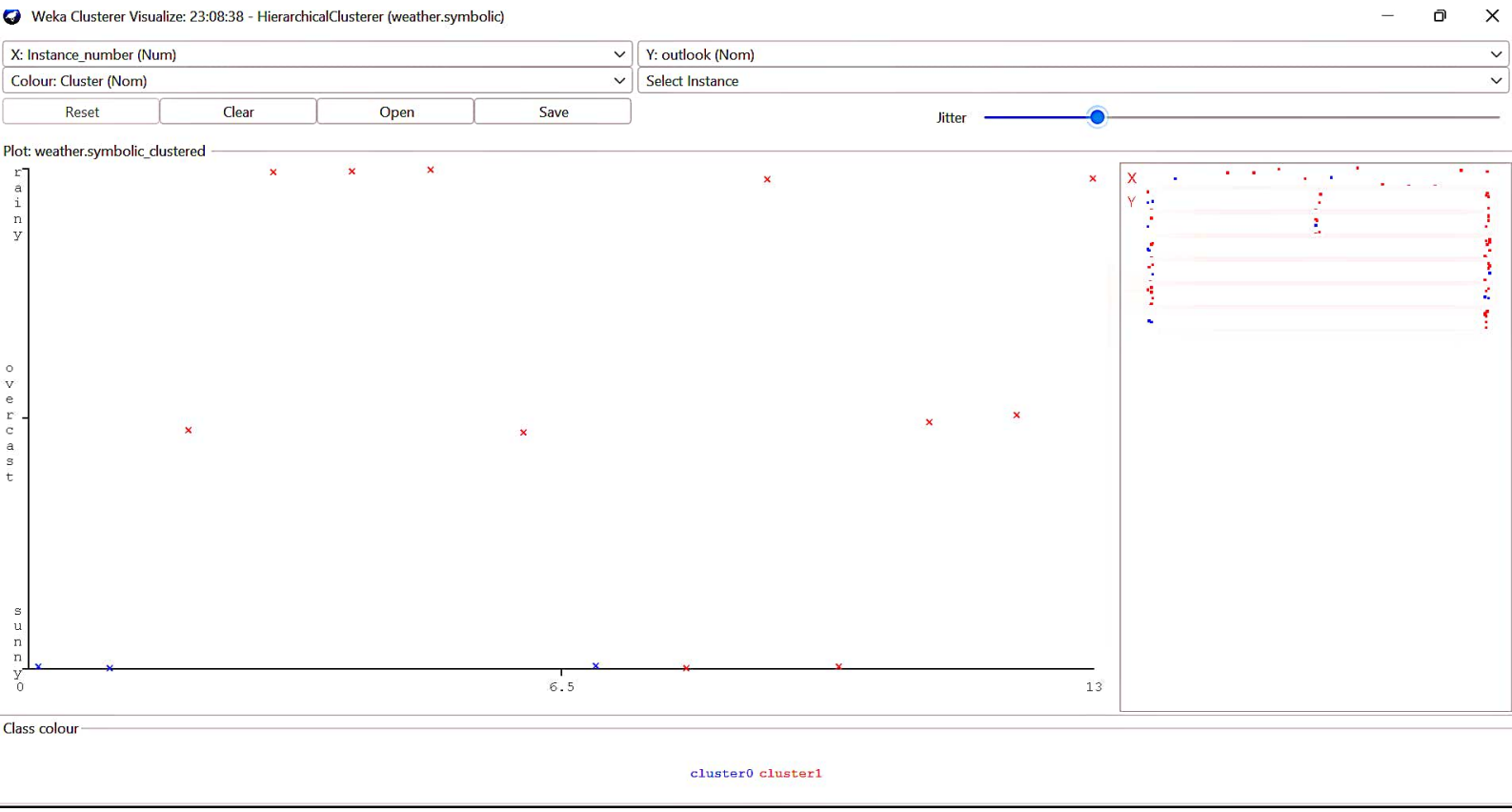
Clusterer output

```
=== Run information ===

Scheme:       weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R first-last"
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
Test mode:    evaluate on training data


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

Cluster 0
((1.0:1,1.0:1):0,1.0:1)

Cluster 1
(((((0.0:1,0.0:1):0.41421,((((0.0:1,0.0:1):0,(0.0:1,0.0:1):0):0.41421,1.0:1.41421):0,0.0:1.41421):0):0,0.0:1.41421):0,0.0:



Time taken to build model (full training data) : 0 seconds

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

Status

OK

Log


Weka Clusterer Visualize: 23:08:38 - HierarchicalClusterer (weather.symbolic)
X: Instance_number (Num)
Y: outlook (Nom)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: weather.symbolic_clustered
rainy
overcast
sunny
0
6.5
13
X
Y
Class colour
cluster0 cluster1