Practical No. 08

Aim: Perform clustering technique on Customer Dataset.

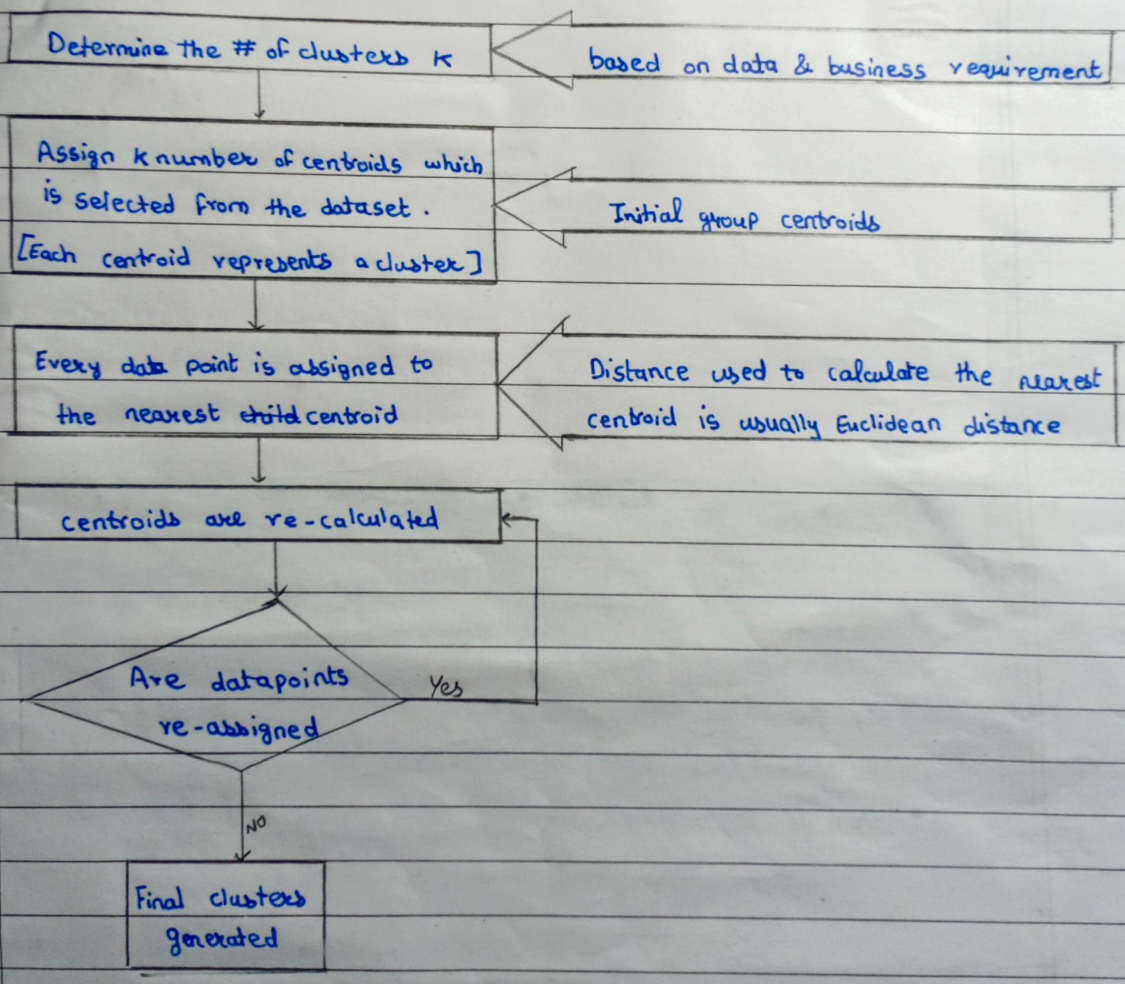
BHAGWATI
PAGE NO. :
DATE :

Practical No. 08

Name of Practical

Aim: Perform Clustering technique on Customer dataset.

Theory: Clustering is an unsupervised machine learning technique, where there are no defined dependent and independent variables. The patterns in the data are used to identify / group similar observations.



Teacher's Signature.....

Practical No. 08

Aim :

Theory :

Weka Explorer

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

Clusterer

Choose **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
- [x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

12:30:13 - SimpleKMeans

Clusterer output

```
kMeans
======

Number of iterations: 7
Within cluster sum of squared errors: 62.1436882815797

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Iris-versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Iris-versicolor

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute               Full Data             0                1
                          (150.0)         (100.0)           (50.0)
=============================================================================
sepallength                5.8433           6.262            5.006
sepalwidth                 3.054            2.872            3.418
petallength                3.7587           4.906            1.464
petalwidth                 1.1987           1.676            0.244
class               Iris-setosa Iris-versicolor     Iris-setosa



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

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

Clustered Instances

0      100 ( 67%)
1       50 ( 33%)
```

Status: OK

Log | x 0

BHAGWATI
PAGE NO. :
DATE :

Name of Practical

K-means clustering:

- K-means clustering is an iterative clustering algorithm where the number of clusters K is predetermined and the algorithm iteratively assigns each data point to one of the K clusters based on the feature similarity.

- The sample dataset used for this example is based on the student data available in ARFF format. This document assumes that appropriate preprocessing has been performed. Steps involved in this Experiment.

Step 1: Run the Weka Explorer and load the data Customer dataset in preprocessing interface.

Step 2: Inorder to perform clustering select the 'cluster' tab in the explorer and clicked on choose button. This step results in a dropdown list of available clustering algorithms.

Step 3: In this case we select 'simple k-means'

Step 4: Next click in text button to the right of the choose button to get pop up window shown in the screenshots. In this window we enter six on the number of clusters and we leave the value of the seed. on as it is. The seed value is used in generating a random number which is used for making the internal assigments of instances of clusters.

Teacher's Signature.....

K-means clustering :

BHAGWATI
PAGE NO. :
DATE :

Name of Practical

Step 5: Once the option has been specified. We run the clustering algorithm there we must make sure they are in the 'cluster mode' panel. The use of training set option is selected and then we click 'start' button. This process and resulting window are shown ~~on~~ in the attached images.

Step 6: The result window shows the centroid of each cluster as well as statistics on the number and the percent of instances assigned to different clusters. Here clusters centroid are means vectors for each clusters. This clusters can be used to characterized ~~th~~ the ~~structure~~ cluster.

Step 7: Another way of understanding characteristics of each cluster through visualisation, we can do this, try right clicking the result set on the result. List panel and selecting the visualize cluster assignments

The mathematics of clustering:

The mathetics-matics behind clustering, in very simple terms involves minimizing the sum of square of distances between the cluster centroids and its associated data points:

$$\text{Minimize} \quad \sum_{J=1}^{K} \sum_{i=1}^{n} (x_{ij} - c_j)^2$$

K = Number of clusters

N = number of data points.

Teacher's Signature.....

Conclusion:

Hence, I performed clustering technique on customer data set.

BHAGWATI
PAGE NO. :
DATE :

Name of Practical

Conclusion:

Hence, I performed clustering technique on customer dataset.

Teacher's Signature.....