

Practical No. 08

Aim: Perform clustering technique on Customer Dataset.

Customer dataset has many fields like customer_id, first_name, last_name, gender, age, etc.
We need to cluster customers into different groups based on their age and gender.

Age ranges from 18 to 60. Gender is either male or female.
Customer IDs are unique and assigned randomly.

We can see that we found three distinct clusters based on gender and age.
Male cluster: Age range 18-40, gender male.

Female cluster: Age range 20-60, gender female.
Mixed cluster: Age range 20-60, gender both.

From the above, we can see that there are three distinct clusters based on gender and age.

Practical No. 08

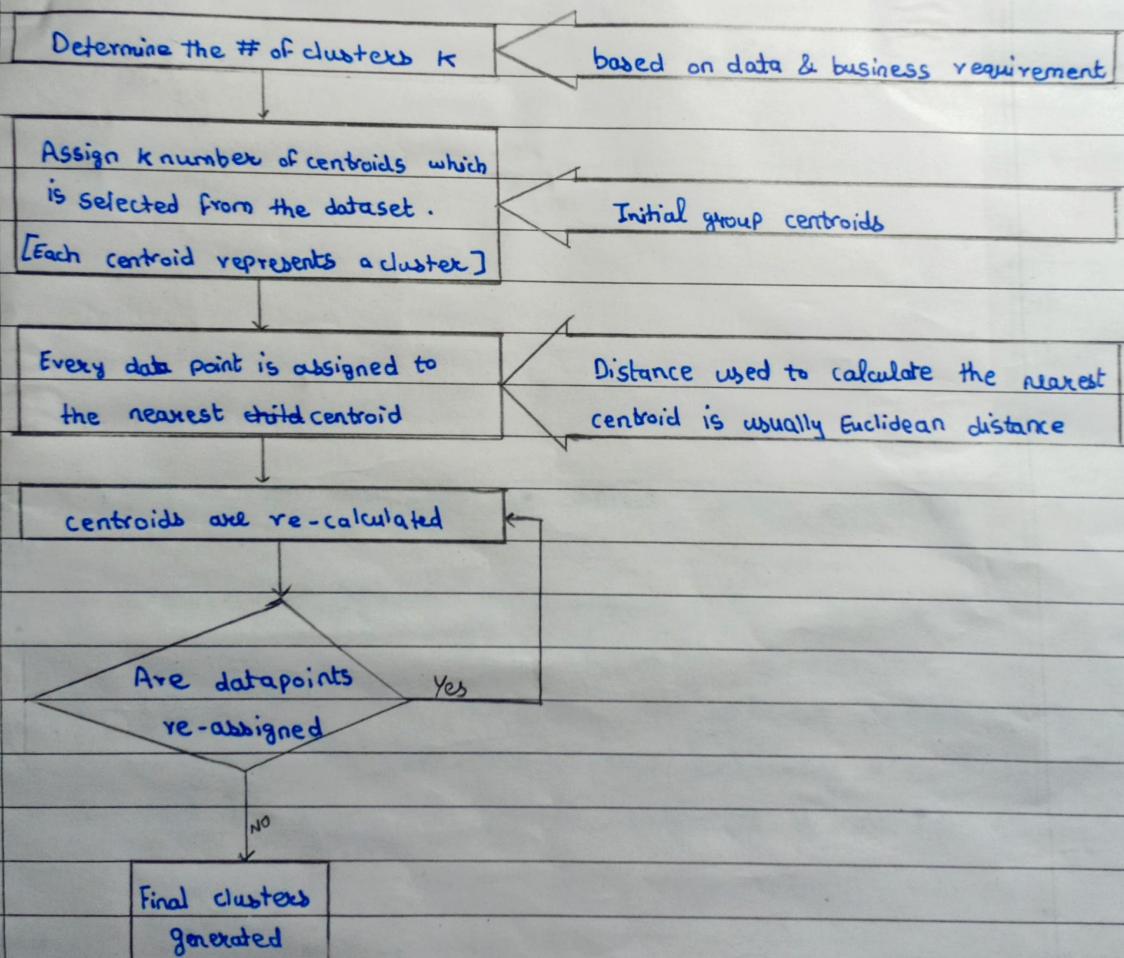
Name of Practical

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

80, 01 Seiten

Imported to email

feindei werden no eigentlich gleichwohl möglich

maA

an der nicht werden eigentlich gleichwohl seien die Regeln no es gleichwohl
und das ist in einem mit anderen Regeln bei technisch keiner
weiteren welche man nicht so sehr

gesetzt

Weka Explorer

- Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose **SimpleKMeans** -inst 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -tl 1-25 -l2 -l0 -N 2 -A "weka.core.EuclideanDistance" -R "first-last" -l 500 -num-slots 1-5 10

Cluster mode

Use training set

Supplied test set

Percentage split

Classes to clusters evaluation (None)

Store clusters for visualization

Ignore attributes

Start Stop

Missing values globally replaced with mean/mode

Result list (right-click for options)

12/2011 SimpleKMeans

Final cluster centroids:		
Attribute	Full Data (150,0)	Cluster# 0 (100,0) Cluster# 1 (50,0)
sepalength	5.0433	6.262 5.006
sepalwidth	3.054	2.872 3.413
petallength	3.7587	4.392 3.855
petalwidth	1.1987	1.676 0.244
class	Iris-setosa Iris-versicolor Iris-virginica	

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

beispiel - er war abhängig

etwa gleich so A

befindet so

wurde kein

berücksichtigt

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: In order to perform clustering select the 'cluster' tab in the explorer and click 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 popup window shown in the screenshots. In this window we enter six on the number of clusters and we leave the value of the seed. as it is. The seed value is used in generating a random number which is used for making the internal assignments of instances of clusters.

Teacher's Signature.....

presently under N.

sittasendo multimediali presentate con le piattaforme esistenti.



ent in diff 'rotates' with twice prior time interval of solvent : \approx 9.76
in either rate int. method seconds no feasible way to do so
- bimolecular process can't be told much about a

'anomalous' fossils see 3402 with ref 15912

notwendig seien, um die Tiere zu schützen. Es ist wichtig, dass die Tiere nicht zu lange ohne Wasser leben. Sie benötigen Wasser, um zu überleben. Wenn sie keinen Wasser trinken, können sie sterben. Es ist wichtig, dass wir die Tiere nicht zu viel Wasser geben, da dies zu einem Überschwund führen kann. Es ist auch wichtig, dass wir die Tiere nicht zu viel Wasser trinken lassen, da dies zu einem Überschwund führen kann.

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 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 mean vectors for each clusters. This clusters can be used to characterize 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 rebuilt list panel and selecting the visualize cluster assignments.

The mathematics of clustering:

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

$$\text{Minimize } \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - c_j)^2$$

k = Number of clusters

n = number of data points.

Teacher's Signature.....

• ni, pietotub bantibd zitomreidtum sitt
perated donatibz to shamps to mre sitt prisimium & tlown' eret shme. for
; atniq tub bantibd sitt has abiorres ztav sitt

details to admin = x

strong club to condemn sin

Hence, I performed clustering technique on customer data set.

Name of Practical

Conclusion:

Hence, I performed clustering technique on customer dataset.

Teacher's Signature.....