

Data warehouse & Data mining lab

Assignment 4a

Submitted by :
Shiv Kumar
2016UIT2563
Semester 8
IT2

Experiment 11 : Write a procedure for Employee data using Make Density Based Cluster Algorithm.

1. Data Loading

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose None Apply Stop

Current relation

Relation: employee Instances: 13 Attributes: 5 Sum of weights: 13

Attributes

All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> eid
2	<input type="checkbox"/> ename
3	<input type="checkbox"/> salary
4	<input type="checkbox"/> exp
5	<input type="checkbox"/> address

Remove

Selected attribute

Name: eid Missing: 0 (0%) Distinct: 13 Type: Numeric Unique: 13 (100%)

Statistic	Value
Minimum	101
Maximum	113
Mean	107
StdDev	3.894

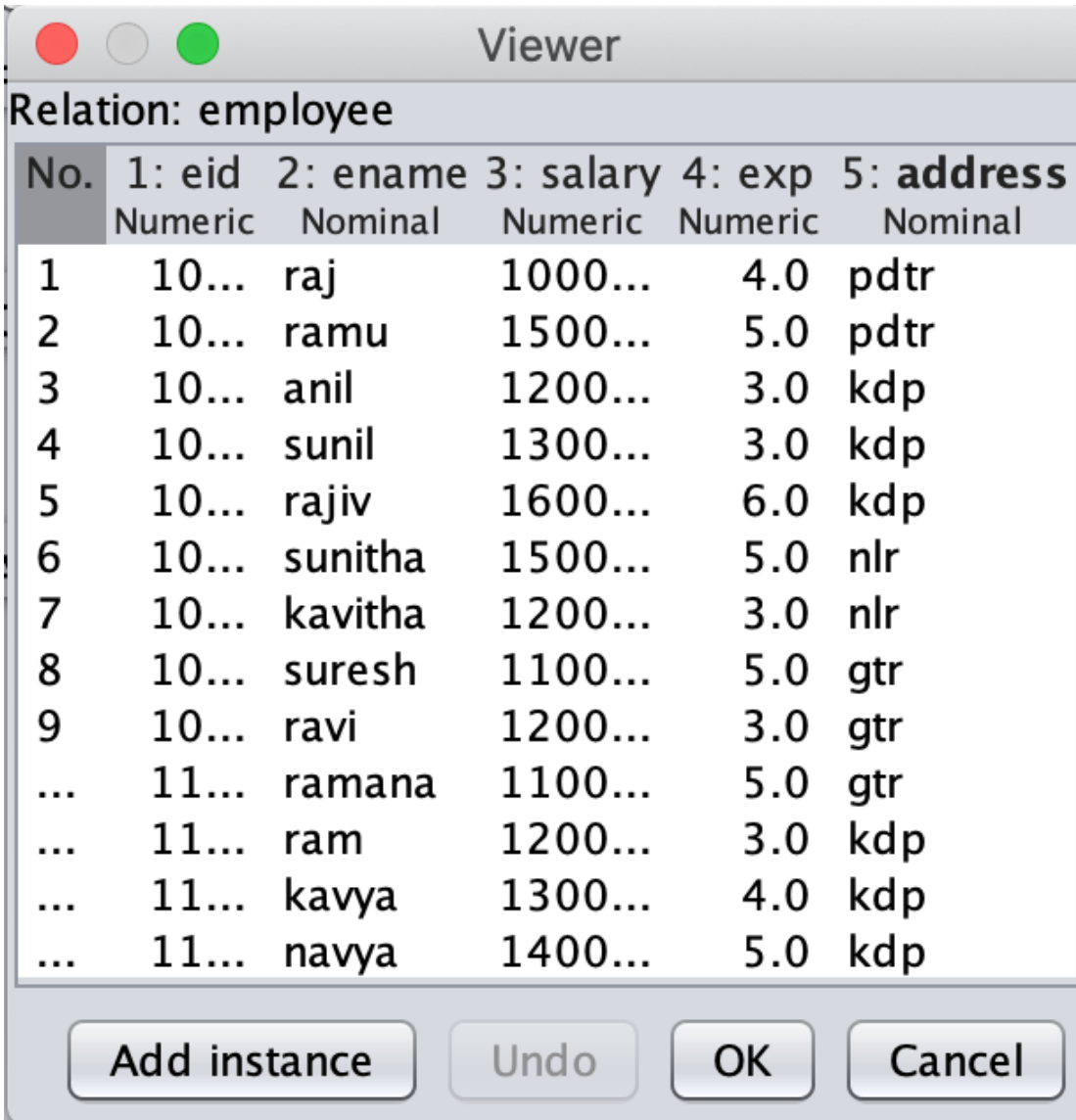
Class: address (Nom) Visualize All

7 6 101 107 113

Status

OK Log x 0

2. Review of Loaded Data



Relation: employee

No.	1: eid Numeric	2: ename Nominal	3: salary Numeric	4: exp Numeric	5: address Nominal
1	10...	raj	1000...	4.0	pdtr
2	10...	ramu	1500...	5.0	pdtr
3	10...	anil	1200...	3.0	kdp
4	10...	sunil	1300...	3.0	kdp
5	10...	rajiv	1600...	6.0	kdp
6	10...	sunitha	1500...	5.0	nlr
7	10...	kavitha	1200...	3.0	nlr
8	10...	suresh	1100...	5.0	gtr
9	10...	ravi	1200...	3.0	gtr
...	11...	ramana	1100...	5.0	gtr
...	11...	ram	1200...	3.0	kdp
...	11...	kavya	1300...	4.0	kdp
...	11...	navya	1400...	5.0	kdp

Add instance Undo OK Cancel

3. Result after Make Density Based Cluster Algorithm.

Weka Explorer

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

Cluster mode

☒ Use training set
☐ Supplied test set Set...
☐ Percentage split % 66
☐ Classes to clusters evaluation
(Nom) address
☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

00:55:48 - MakeDensityBasedClusterer

Clusterer output

```
=== Run information ===  
  
Scheme:      weka.clusterers.MakeDensityBasedClusterer -M 1.0E-6 -W weka.clusterers.SimpleKMeans  
Relation:    employee  
Instances:   13  
Attributes:  5  
             eid  
             ename  
             salary  
             exp  
             address  
  
Test mode:   evaluate on training data  
  
=== Clustering model (full training set) ===  
  
MakeDensityBasedClusterer:  
  
Wrapped clusterer:  
kMeans  
=====
```

Number of iterations: 2
Within cluster sum of squared errors: 19.3315972222222

Initial starting points (random):
Cluster 0: 103,anil,12000,3,kdp
Cluster 1: 101,raj,10000,4,pdtr

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (13.0)	Cluster#	
		0 (9.0)	1 (4.0)
eid	107	107.7778	105.25
ename	raj	anil	raj
salary	12769.2308	13222.2222	11750
exp	4.1538	3.8889	4.75
address	kdn	kdn	pdtr

Status

OK Log x 0

Final cluster centroids:

Attribute	Full Data (13.0)	Cluster#	
		0 (9.0)	1 (4.0)
eid	107	107.7778	105.25
ename	raj	anil	raj
salary	12769.2308	13222.2222	11750
exp	4.1538	3.8889	4.75
address	kdp	kdp	pdtr

Fitted estimators (with ML estimates of variance):

Cluster: 0 Prior probability: 0.6667

Attribute: eid

Normal Distribution. Mean = 107.7778 StdDev = 3.4247

Attribute: ename

Discrete Estimator. Counts = 1 1 2 2 2 2 2 1 2 1 2 2 2 (Total = 22)

Attribute: salary

Normal Distribution. Mean = 13222.2222 StdDev = 1396.645

Attribute: exp

Normal Distribution. Mean = 3.8889 StdDev = 1.0999

Attribute: address

Discrete Estimator. Counts = 1 7 3 2 (Total = 13)

Cluster: 1 Prior probability: 0.3333

Attribute: eid

Normal Distribution. Mean = 105.25 StdDev = 3.8324

Attribute: ename

Discrete Estimator. Counts = 2 2 1 1 1 1 1 2 1 2 1 1 1 (Total = 17)

Attribute: salary

Normal Distribution. Mean = 11750 StdDev = 1920.2864

Attribute: exp

Normal Distribution. Mean = 4.75 StdDev = 0.433

Attribute: address

Discrete Estimator. Counts = 3 1 1 3 (Total = 8)

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0 9 (69%)

1 4 (31%)

Log likelihood: -16.52967