

EX.No: 07

Date :

DATASEGMENTATION BY EXPECTATION MAXIMISATION ALGORITHM THROUGH WEKA

CLUSTER (EM)

=== Run information ===

Scheme:weka.clusterers.EM -I 100 -N -1 -M 1.0E-6 -
S 100

Relation: crime

Instances: 47

Attributes: 8

Year

Population

Murder

Rape



Robbery

Assault

Burglary

CarTheft

Test mode:evaluate on training data

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

EM

==

Number of clusters selected by cross validation: 5

	Cluster			
Attribute	0	1	2	3
4	(0.26)	(0.13)	(0.36)	(0.11)
	(0.15)			



=====

=====

=

Year

mean	2005.5	1990.5	1980.3529
1997	1968		
std. dev.	3.4521	1.7078	5.625
1.4142	2		

Population

mean	19289794.1667	18035409.6788	
17881041.974	18165920.2042	18240390.4563	
std. dev.	160346.5501	101493.8891	
248964.7389	24977.5465	110418.0323	

Murder

mean	874.9167	2413.834	1973.7646
1164.6	1212.8533		
std. dev.	62.0396	140.5813	128.9
251.188	324.551		

Rape



mean	3250.4167	5222.3324	5187.2304
3989	2707.5683		
std. dev.	441.2481	161.8788	377.3213
258.849	290.3994		

Robbery

mean	33396.75	106069.1992	
93395.4876	56670.798	57417.4533	
std. dev.	3707.4057	5447.0007	10310.3002
9992.3401	24323.9592		

Assault

mean	48221.3333	89751.8278	
61470.1658	64743.7988	34452.236	
std. dev.	5417.1056	2289.2575	10316.9741
5207.2629	5032.6828		

Burglary

mean	71016.0833	202959.7771	
269802.6448	118546.7964	234185.4531	
std. dev.	7280.2331	12052.4815	
51879.9057	18666.3922	32520.6333	



CarTheft

mean	34908.9167	169109.0575	
	121974.5885	79724.9971	97171.7993
std. dev.	11617.2257	13034.3403	11011.47
	15625.6976	26425.6258	

Time taken to build model (full training data) : 0.21 seconds

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

Clustered Instances

0	12 (26%)
1	6 (13%)
2	17 (36%)
3	5 (11%)
4	7 (15%)



Log likelihood: -73.74065

SIMPPL K MEANS

=== Run information ===

Scheme:weka.clusterers.SimpleKMeans -N 2 -A
"weka.core.EuclideanDistance -R first-last" -I 500 -S
10

Relation: crime

Instances: 47

Attributes: 8

Year

Population

Murder

Rape

Robbery

Assault



Burglary

CarTheft

Test mode:evaluate on training data

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

kMeans

=====

Number of iterations: 4

Within cluster sum of squared errors:

13.42935674288335

Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#	
Attribute	Full Data	0	1
	(47)	(20)	(27)
=====			



=====

Year	1988	1996.1	1982
Population	18344255.6596	18847556.55	17971440.1852
Murder	1549.9787	933.55	2006.5926
Rape	4200.4255	3230.55	4918.8519
Robbery	70429.2979	38498.1	94082.037
Assault	58022.234	47541.85	65785.4815
Burglary	189119.8298	107400.5	249652.6667
CarTheft	97573.5532	51320.4	131835.1481

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

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

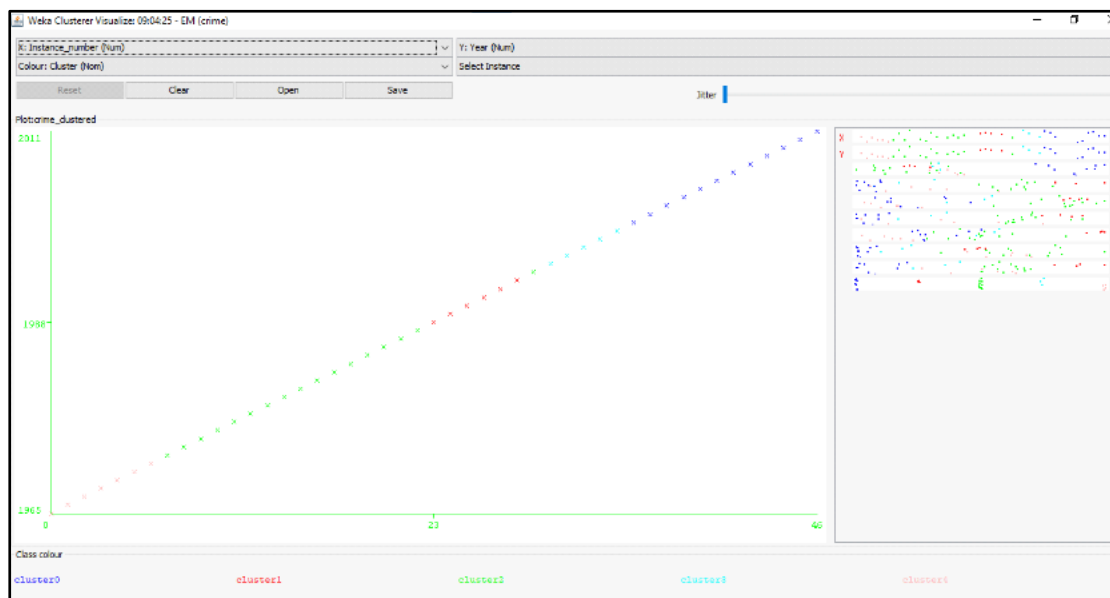


Clustered Instances

0 20 (43%)

1 27 (57%)

EM CLUSTER VISUALISATIONS



SIMPLE K MEANS CLUSTER VISUALISATIONS

