

Algorithm: K-means

Dataset: Bank Telemarketing data

Number of attributes: 17 Number of instances: 100

Distance Measure	No of Clusters(k)	Cluster Sizes(no of data points/cluster)
Euclidean	2	96,4
Euclidean	3	4,82,14
Euclidean	4	66,10,3,21
Euclidean	5	12,58,7,4,19
Manhattan	2	4,96
Manhattan	3	4,82,14
Manhattan	4	20,3,70,7
Manhattan	5	9,18,59,4,10
Least Squares	2	8,92
Least Squares	3	4,82,14
Least Squares	4	4,64,22,10
Least Squares	5	10,2,22,2,64

No of Clusters(k)	Distance Measure	Cluster Sizes(no of data points /cluster)
2	Euclidean	96,4
2	Manhattan	4,96
2	Least Squares	8,92
3	Euclidean	4,82,14
3	Manhattan	4,82,14
3	Least Squares	4,82,14
4	Euclidean	66,10,3,21
4	Manhattan	20,3,70,7
4	Least Squares	4,64,22,10
5	Euclidean	12,58,7,4,19
5	Manhattan	9,18,59,4,10
5	Least Squares	10,2,22,2,64

Validation results: Phase I

K clusters with specified methods*	Average Silhouette Width	
	Independent	Ensemble [#]
2 - Euclidean	0.84	0.83
2 - Manhattan	0.83	0.84
3 – Euclidean	0.72	0.69
3 – Manhattan	0.69	0.72
4 – Euclidean	0.61	0.58
4 – Manhattan	0.58	0.62
5 – Euclidean	0.50	0.37
5 – Manhattan	0.47	0.59

2 - CLARA – Euclidean	0.72	0.71
2 - CLARA - Manhattan	0.68	0.72

* Default clustering algorithm is k-means

If the initial clustering is done with Euclidean, boosting is done with Manhattan and vice-versa.