CIS 635 Data Mining Name Sajal Shrestha

Homework 8

1. normalized values

	x_1	x_2
101	0.2	0.45
102	0.5	0.0
103	0.4	1.0
104	0.0	0.24
105	1.0	0.61
106	0.7	0.10

2. manual assignments: fill in the tables for the calculation of centroids and assignment of clusters.

means	iteration 1		iteration 2	
cluster	x_1	x_2	x_1	x_2
1	9.6	208	18.076	82.668
2	9.6	279.6	82.761	32.212

	first assignment			second assignment		
id	dist. to cluster 1	dist. to cluster 2	cluster assign.	dist. to cluster 1	dist to cluster 2	cluster assign.
101	43.078	28.717	cluster 2	59.462	39.836	cluster 2
102	30.002	101.6008	cluster 1	22.376	87.167	cluster 1
103	134.001	62.492	cluster 2	117.683	59.473	cluster 2
104	11.007	61.771	cluster 1	22.059	77.603	cluster 1
105	71.205	5.433	cluster 2	93.743	29.165	cluster 2
106	13.219	84.634	cluster 1	5.239	87.086	cluster 1

3. Agglomerative clustering
   a) distance matrix

	c1	c2	c3	c4	c5	c6
c1	0	73.06162	91.02198	33.06055	29.12044	56.22277
c2	73.06162	0	164.0031	40.31129	101.1237	17.11724
c3	91.02198	164.0031	0	124.0645	63.28507	147.0306
c4	33.06055	40.31129	124.0645	0	61.81424	24.04163
c5	29.12044	101.1237	63.28507	61.81424	0	84.05355
c6	56.22277	17.11724	147.0306	24.04163	84.05355	0

   b) answers:

   ◦ what two clusters would be joined next: c2 and c6

   ◦ what would be the new distance: 147.030

4. R analysis
   a) sizes 664 1237 1256 1225 618

   b) descriptions

   - cluster 1 The people belonging to this cluster are mature. They generally tend to have significant expenditure in media and housing. However, their expense on pets is small.

   - cluster 2 The people in this cluster spend more money on their pets. On the other hand, they seem to have low expenditure on media.

   - cluster 3 The people associated with this cluster are comparatively younger. They seem to have average expenses across multiple categories.

   - cluster 4 The people belonging in this group spend most of their money on housing and pets. Likewise, their food expenses are somewhat in the upper range.

   - cluster 5 The people in this cluster spend the most on food and transport as compared to other clusters. However, their housing expenses are low.