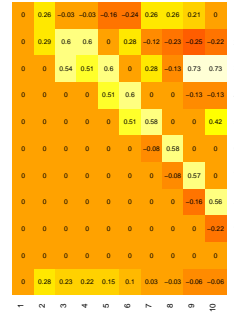
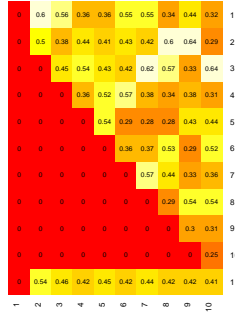


(a) Hierarchical Clustering

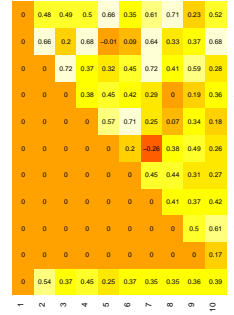


(b) Agglomerative Clustering

Figure 1: Average Silhouette Plots



(a) K-means Clustering



(b) Spectral Clustering

Figure 2: Average Silhouette Plots

Homework 3 Partial Solutions

Part 1: For this problem, I was looked for the basic clustering and some comparison across a range of k . I tried k from 2 through 10. The average silhouette plots are included below. The average silhouette scores for each plot are shown below. In addition, the last row shows the average over all clusters. In most cases, $k = 2$, seems to give the best results.

Part 2: The confusion matrix for clusters versus area and clusters versus region are show below. We use spectral clustering as the performance was pretty good for that. In general, it seems like the clustering is decent when compared to the real data, but there isn't 100% accuracy. In the first case, 279 from Area 1 gets put in cluster 1, while in the second table, 223 from Area 1 gets put in cluster 3 and so on.

Cluster/Area	1	2	3
1	279	98	4
2	44	0	147

Cluster/Area	1	2	3
1	56	71	10
2	44	0	141
3	223	27	0

Cluster/Area	1	2	3
1	195	18	0
2	3	80	12
3	87	0	11
4	38	0	128

Cluster/Region	1	2	3	4	5	6	7	8	9
1	0	0	1	0	65	0	0	0	0
2	12	0	0	1	0	0	23	2	51
3	0	0	11	1	0	0	0	0	0
4	0	0	1	0	0	33	0	0	0
5	13	55	15	26	0	0	27	48	0
6	0	1	178	8	0	0	0	0	0

Cluster/Region	1	2	3	4	5	6	7	8	9
1	1	51	15	13	0	0	2	1	0
2	0	0	1	0	65	0	0	2	0
3	0	0	29	1	0	0	0	0	0
4	0	1	157	8	0	0	0	0	0
5	0	0	4	0	0	33	0	0	0
6	14	4	0	14	0	0	35	46	0
7	10	0	0	0	0	0	13	1	51

Part 3: Finally we used spectral clustering for 2 centers and performed random forest, lda and qda classification. The error rates, based on 10 fold cross-validation, were 0.9930127, 0.9720811 and 0.9790684 respectively. This isn't much higher than the previous weeks solutions.

Cluster/Region	1	2	3	4	5	6	7	8	9
1	15	4	0	12	0	0	34	47	0
2	0	0	149	4	0	0	0	0	0
3	1	51	15	13	0	0	3	1	0
4	0	1	28	4	0	0	0	0	0
5	0	0	1	0	65	0	0	1	0
6	9	0	0	0	0	0	13	1	51
7	0	0	1	0	0	33	0	0	0
8	0	0	12	3	0	0	0	0	0

Cluster/Region	1	2	3	4	5	6	7	8	9
1	0	0	54	1	0	0	0	0	0
2	10	0	0	1	0	0	11	1	51
3	0	0	41	2	0	0	0	0	0
4	12	5	0	9	0	0	37	8	0
5	0	1	88	6	0	0	0	0	0
6	0	0	10	0	0	33	0	0	0
7	2	0	0	4	0	0	0	38	0
8	0	0	1	1	65	0	0	2	0
9	1	50	12	12	0	0	2	1	0

Cluster/Region	1	2	3	4	5	6	7	8	9
1	0	0	1	0	0	33	0	0	0
2	14	4	0	12	0	0	34	9	0
3	0	0	12	3	0	0	0	0	0
4	3	0	0	0	0	0	4	0	51
5	1	51	15	13	0	0	3	1	0
6	0	0	1	0	65	0	0	1	0
7	0	1	138	8	0	0	0	0	0
8	0	0	0	0	0	0	0	38	0
9	7	0	0	0	0	0	9	1	0
10	0	0	39	0	0	0	0	0	0