Andrew
Curtis

Assignment 8

1] a) cluster assignment

iter	0.090	0.172	0.310	0.429	0.640	0.642	0.851	Centroid Locations A	B	C
0								0.15	0.25	0.90
1	A	A	B	B	C	C	C	0.131	0.370	0.711
2	A	A	B	B	C	C	C	0.131	0.370	0.711

* The algorithm converged and did not change between iter 1 and 2.

b)

iter	0.090	0.172	0.310	0.429	0.640	0.642	0.851	Centroid Locations A	B	C
0								0.10	0.45	0.90
1	A	A	B	B	B	B	C	0.131	0.505	0.851
2	A	A	A	B	B	B	C	0.191	0.570	0.851

* while the 3rd iteration has not been presented, it would be converged after the 2nd iteration. Interestingly, the assignment and locations are different than in A).

c) compute SSE for a)

$$SSE = (0.09-0.131)^2 + (0.172-0.131)^2 + (0.310-0.370)^2 + (0.429-0.370)^2 + (0.64-0.711)^2 + (0.642-0.711)^2 + (0.851-0.711)^2$$

$$SSE = 0.0398$$

d) compute SSE for b)

$$SSE = (0.09-0.191)^2 + (0.172-0.191)^2 + (0.310-0.191)^2 + (0.429-0.570)^2 + (0.640-0.570)^2 + (0.642-0.570)^2 + (0.851-0.851)^2$$

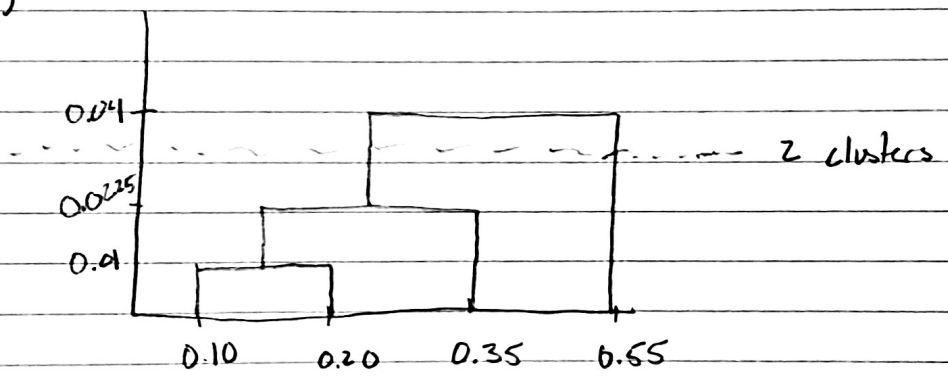
$$SSE = 0.0547$$

* Solution a) had lower SSE and was thus better.

2| a)

	0.10	0.20	0.35	0.55
0.10	0	0.01	0.0625	0.2025
0.20	0.01	0	0.0225	0.1225
0.35	0.0625	0.0225	0	0.04
0.55	0.2025	0.1225	0.04	0

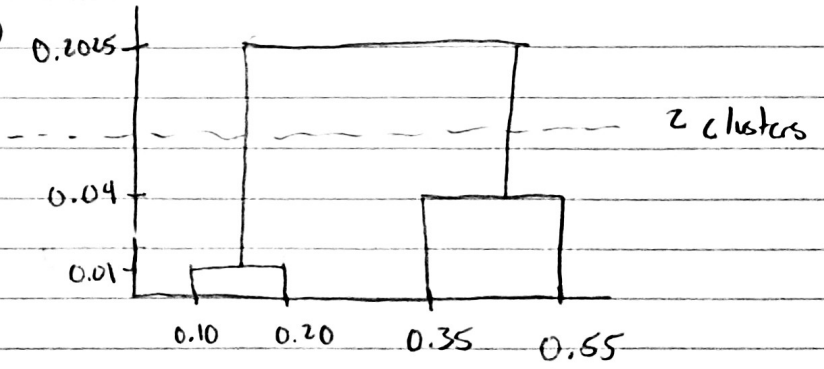
b)



$SSE = (0.10 - 0.217)^2 + (0.20 - 0.217)^2 + (0.35 - 0.217)^2 + (0.55 - 0.55)^2$

$SSE = 0.032$

c)



$SSE = (0.10 - 0.15)^2 + (0.20 - 0.15)^2 + (0.35 - 0.45)^2 + (0.55 - 0.45)^2$

$SSE = 0.025$

2) d)

iter	0.10	0.20	0.35	0.55	Centroids A	Centroids B
0					0.15	0.5
1	A	A	B	B	0.15	0.45
2	A	A	B	B	0.15	0.45

$$SSE = (0.10-0.15)^2 + (0.20-0.15)^2 + (0.35-0.45)^2 + (0.55-0.45)^2$$

$$SSE = 0.025$$

* the K-means and Max algorithm clustering produced the same clusters and share the lowest SSE value.

the min clustering algo had 3-point and 1-point clusters with a higher overall SSE

3)

Algo A $ARI = \dfrac{\binom{100}{2}(10+15) - ((10+35)(10+40) + (40+15)(35+15))}{\binom{100}{2}^2 - ((10+35)(10+40) + (40+15)(35+15))}$

Algo A $ARI = \dfrac{118750}{24497500} = 0.00485$

Algo B $ARI = \dfrac{\binom{100}{2}(35+10) - ((35+40)(35+15) + (15+10)(40+10))}{\binom{100}{2}^2 - ((35+40)(35+15) + (15+10)(40+10))}$

Algo B $ARI = \dfrac{217750}{24497500} = 0.00889$

* used adjusted random index (ARI), Solution B has the higher ARI value of 0.00889 indicating a better solution.