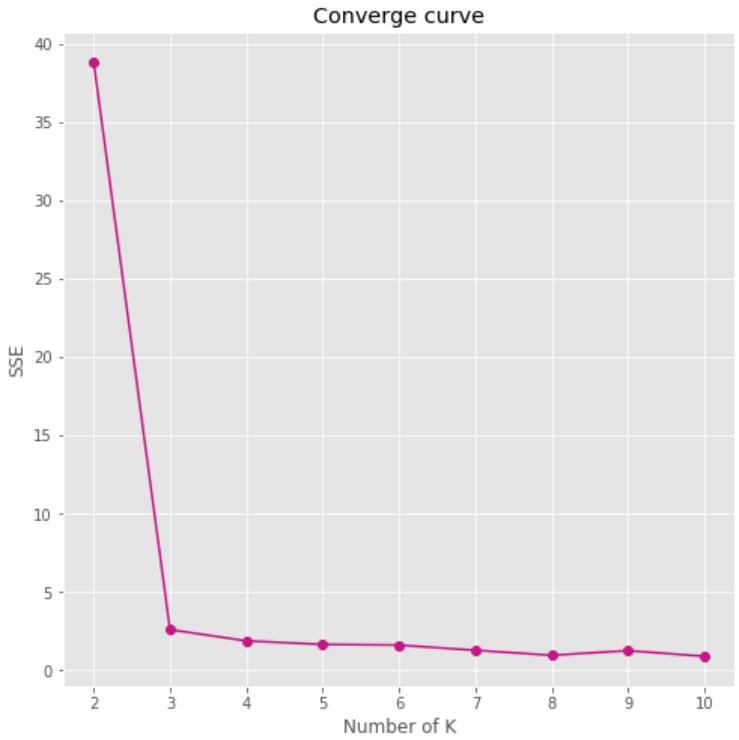
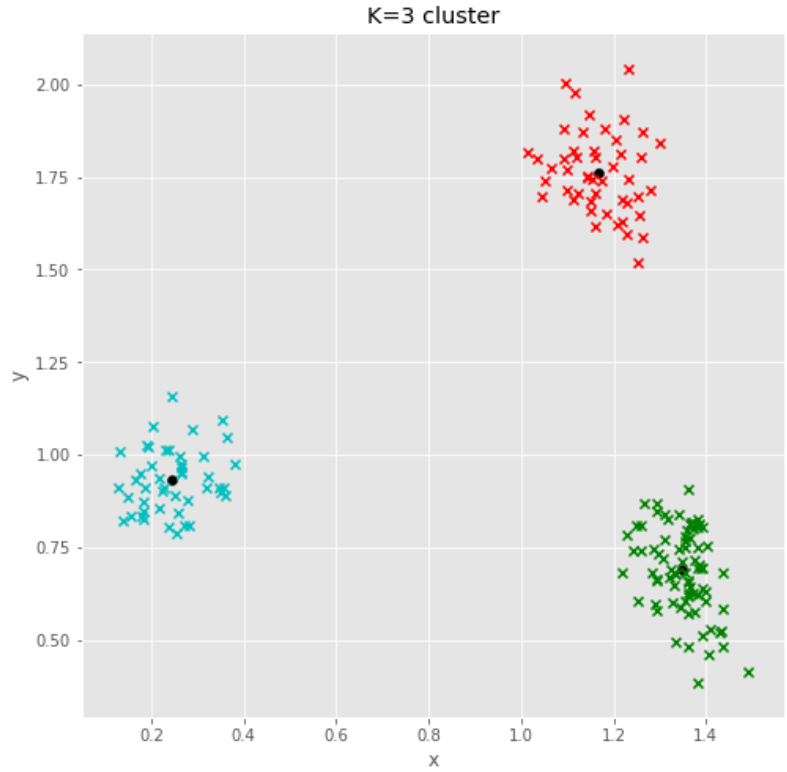
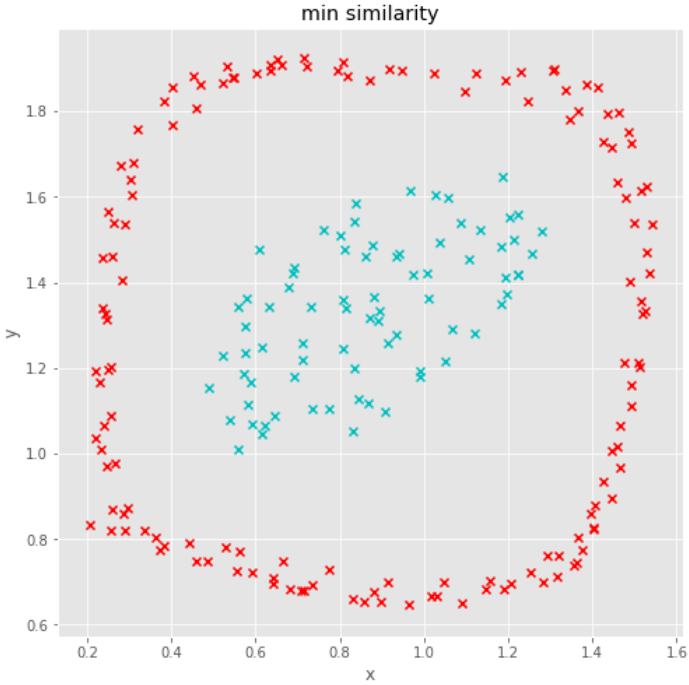
1.(1)

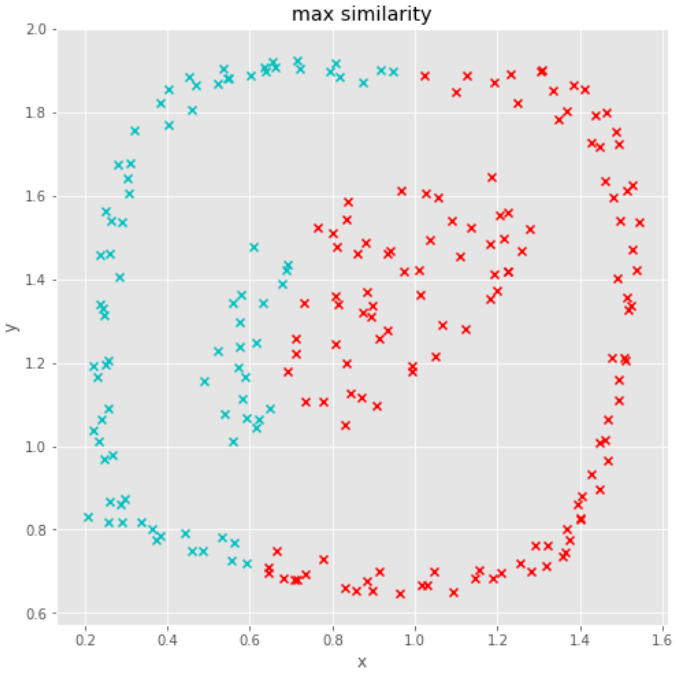

Converge curve
SSE
Number of K
0
5
10
15
20
25
30
35
40
2
3
4
5
6
7
8
9
10

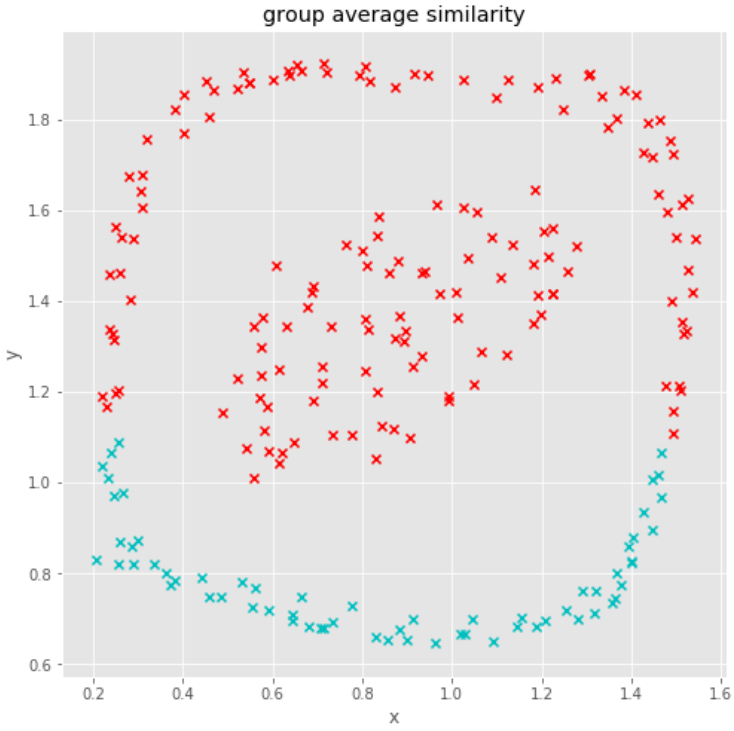
(2)

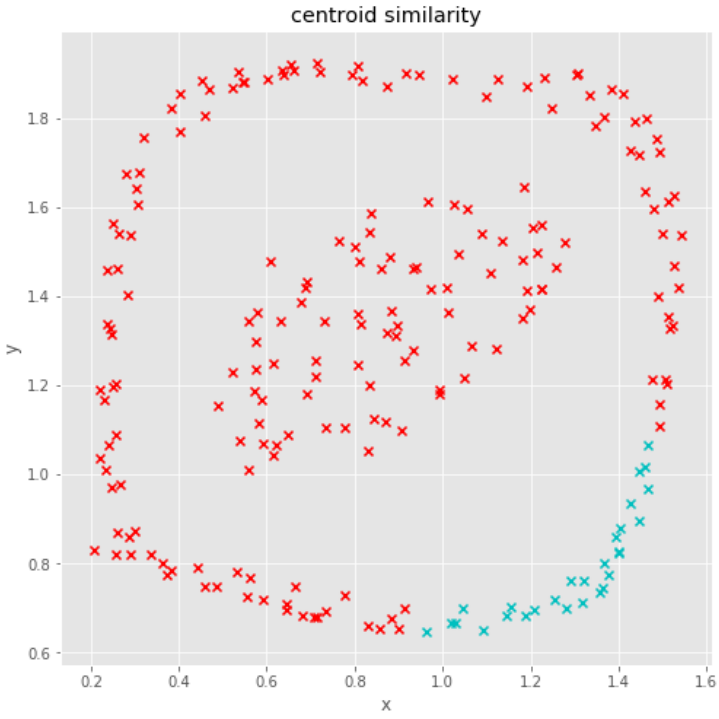

K=3 cluster
y
x
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.2
0.4
0.6
0.8
1.0
1.2
1.4

2.


min similarity
y
x
0.6
0.8
1.0
1.2
1.4
1.6
1.8
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6


max similarity
y
x
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6




Result:

Min similarity give us the best result for this kind of dataset. Max tends to break the large clusters that is the reason we get separate inner and outer cluster. Min algorithm handle non-elliptical shapes well but very sensitive to outliner and noise. Group average is tradeoff between min and max. but the limitation is Biased towards globular clusters. Centroid have strength in global shaped dataset. All of those have pros and cons.