

Sharice Mayer

CS 445 Spring 2020

Programming #3 Report - K-mean and Fuzzy C-Means

6/01/20

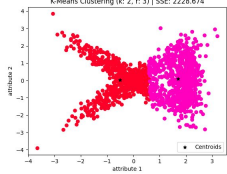
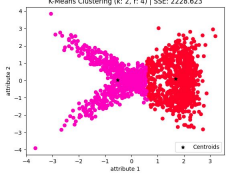
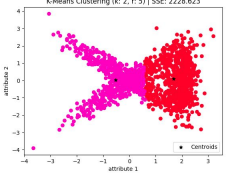
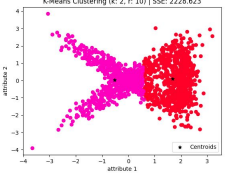
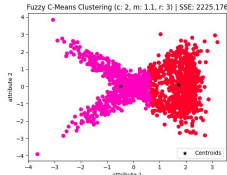
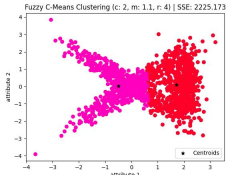
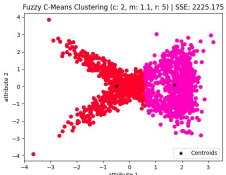
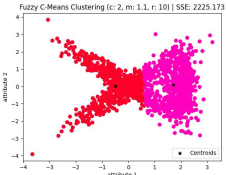
Program Description:

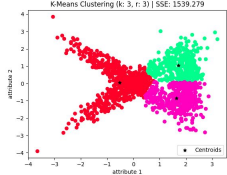
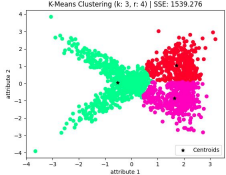
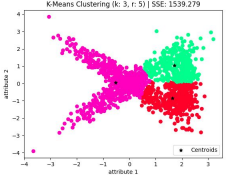
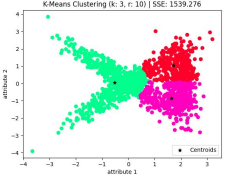
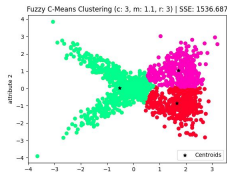
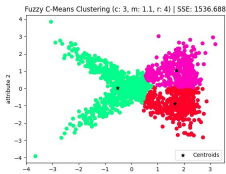
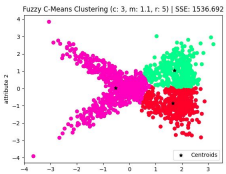
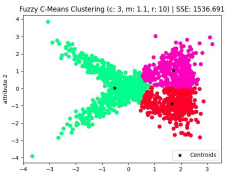
- The dataset used for this program was given, and included 1500 data points of 2d data (for ease of visualization) simulated from 3 Gaussians, with considerable overlap.
- There are 500 points from each Gaussian, ordered together in the file
- K-Means Clustering Algorithm was used to classify the data, as well as Fuzzy C-Means for comparison
- Both algorithms were tested for $K = 2, 3, 4, 5, 6, 7, 8$, and 9
- Each K value was run for 3, 4, 5, 6, 7, 8, 9, and 10 separate runs
- ** For brevity, only plots from $R = 3, 4, 5$, and 10 were included in this report
- The 'm' parameter for Fuzzy C-Means Clustering used in this report was 1.1

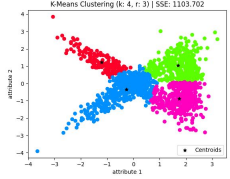
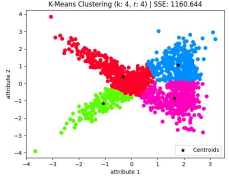
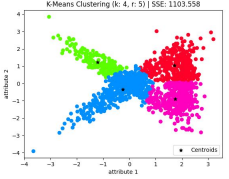
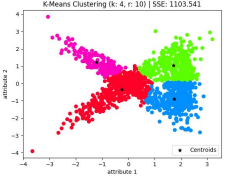
Results:

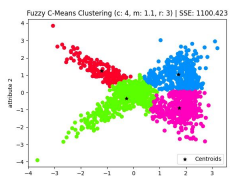
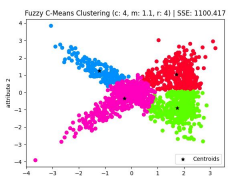
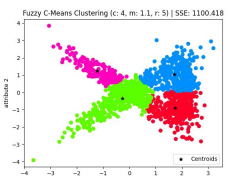
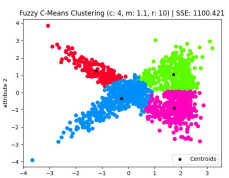
K-Means clustering took a fraction of the time that the Fuzzy C-Means algorithm took, which became increasingly more evident as the algorithms ran for greater numbers of clusters. For both algorithms, as the number of clusters increased, the SSE decreased, which was expected given that more clusters decrease the maximum distance a particular point must travel to get to the closest centroid. Also, the SSE variance between identical k -values for varying runs increased as k did for both algorithms, and the clustering structure of C-Means became less consistent between varying runs. Both algorithms yield similar results during clustering for the same k -values, though K-Means had slightly higher SSE in each test.

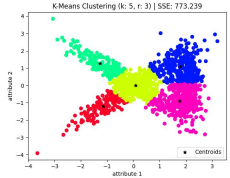
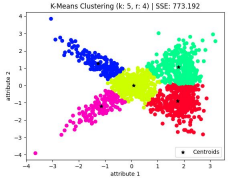
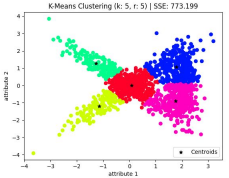
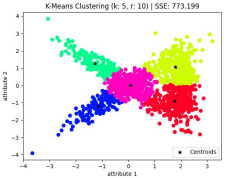
PLOTS:

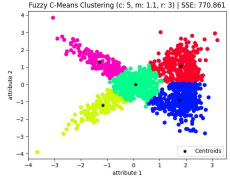
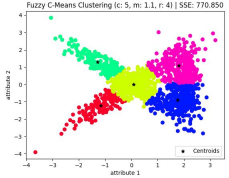
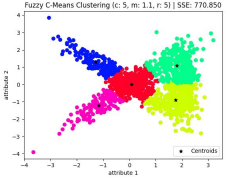
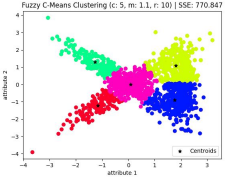
K = 2	R = 3	R = 4	R = 5	R = 10
SSE	2228.67400613	2228.62330329	2228.62330329	2228.62330329
K-Means				
K = 2	R = 3	R = 4	R = 5	R = 10
SSE	2225.17564758	2225.17288786	2225.17486522	2225.17315739
C-Means				

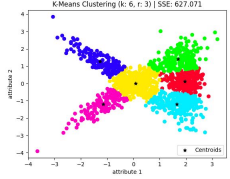
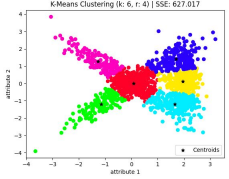
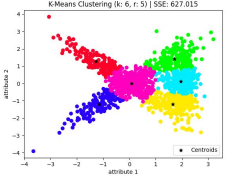
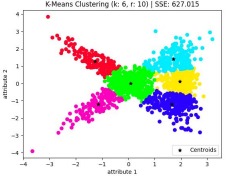
K = 3	R = 3	R = 4	R = 5	R = 10
SSE	1539.27940111	1539.27608880	1539.27940111	1539.27608880
K-Means				
K = 3	R = 3	R = 4	R = 5	R = 10
SSE	1536.68660839	1536.68836795	1536.69212067	1536.69144367
C-Means				

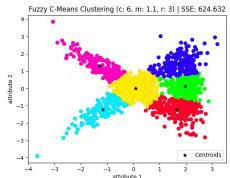
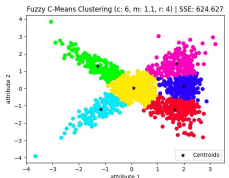
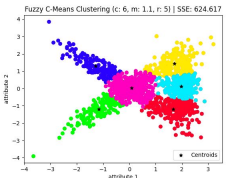
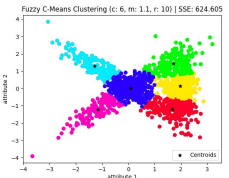
K = 4	R = 3	R = 4	R = 5	R = 10
SSE	1103.70203699	1160.64414530	1103.55762535	1103.54128133
K-Means				

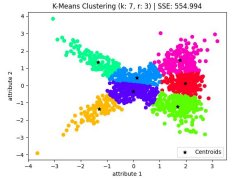
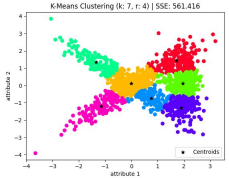
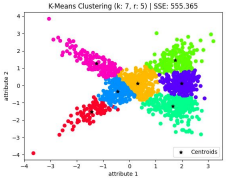
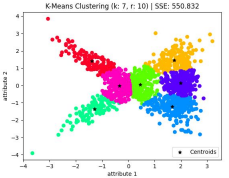
K = 4	R = 3	R = 4	R = 5	R = 10
SSE	1100.42294394	1100.41668044	1100.41802473	1100.42087124
C-Means				

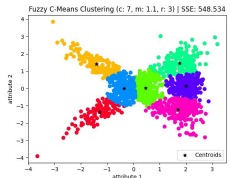
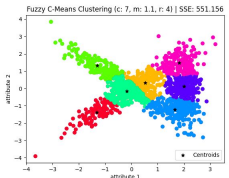
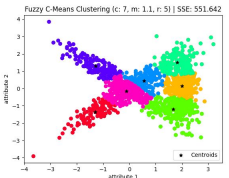
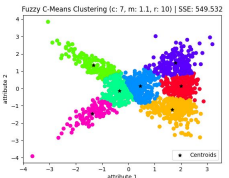
K = 5	R = 3	R = 4	R = 5	R = 10
SSE	773.23923763	773.19176363	773.19889256	773.19889256
K-Means				

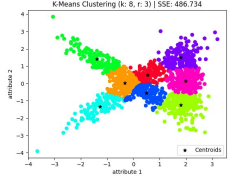
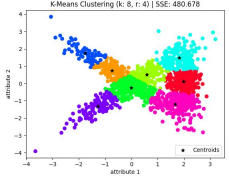
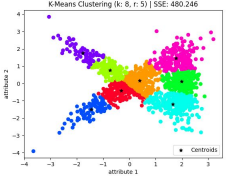
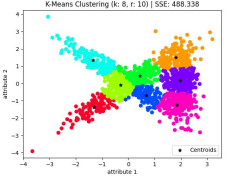
K = 5	R = 3	R = 4	R = 5	R = 10
SSE	770.86081177	770.85036529	770.85048054	770.84733408
C-Means				

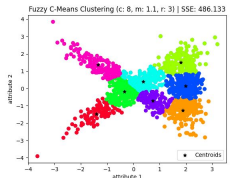
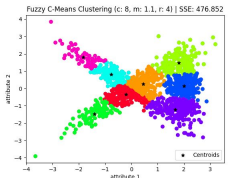
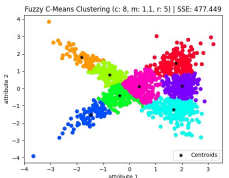
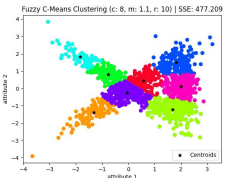
K = 6	R = 3	R = 4	R = 5	R = 10
SSE	627.07079778	627.01725174	627.01540260	627.01540260
K-Means				

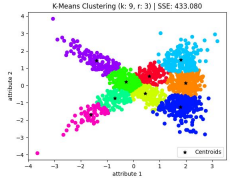
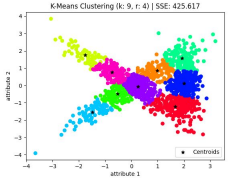
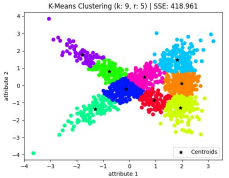
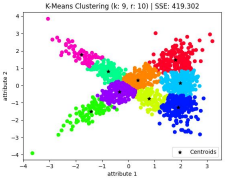
K = 6	R = 3	R = 4	R = 5	R = 10
SSE	624.63242723	624.62717547	624.61702731	624.60541168
C-Means				

K = 7	R = 3	R = 4	R = 5	R = 10
SSE	554.99363007	561.41611684	555.36501060	550.83158884
K-Means				

K = 7	R = 3	R = 4	R = 5	R = 10
SSE	548.53399733	551.15620980	551.64157382	549.53175658
C-Means				

K = 8	R = 3	R = 4	R = 5	R = 10
SSE	486.73435330	480.67830257	480.24582736	488.33771797
K-Means				

K = 8	R = 3	R = 4	R = 5	R = 10
SSE	486.13294678	476.85155467	477.44916000	477.20912005
C-Means				

K = 9	R = 3	R = 4	R = 5	R = 10
SSE	433.08031480	425.61738040	418.96145999	419.30241581
K-Means				

K = 9	R = 3	R = 4	R = 5	R = 10
SSE	416.18131815	416.94143682	422.96165542	416.22197737
C-Means	