Sharice Mayer
CS 445 Spring 2020
Programming #3 Report - K-mean and Fuzzy C-Means
6/01/20

Program Description:

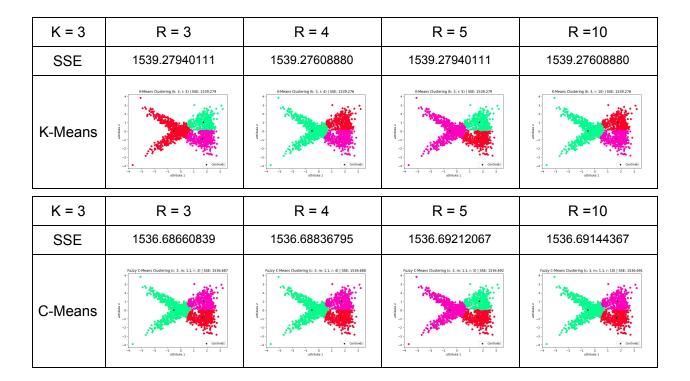
- The dataset used for this program was given, and included 1500 data points of 2d data (for ease of visualization) simulated from 3 Gaussians, with considerable overlap.
- There are 500 points from each Gaussian, ordered together in the file
- K-Means Clustering Algorithm was used to classify the data, 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	KAMeans Clustering (b: 2, r: 3)) SSE 2228.674	Kilmans Clustering (In: 2, rr. 4) 1555: 2228.623	Kilmans Clustering (b: 2, r: 57) ISSE: 2228.623	EMeans Clustering (s. 2, r. 10) 1555: 2228.623
K = 2	R = 3	R = 4	R = 5	R =10
K = 2 SSE	R = 3 2225.17564758	R = 4 2225.17288786	R = 5 2225.17486522	R =10 2225.17315739



K = 4	R = 3	R = 4	R = 5	R =10
SSE	1103.70203699	1160.64414530	1103.55762535	1103.54128133
K-Means	Kolheans Clustering (b. 4, e. 3) SSE: 1103.702	K-Means Clustering (b. 4, r. 4) SSE: 1160:644	Kolhains Clustering (b. 4, n. 5) SSE: 1103:558	K. Orleans Clustering (tr. d., [r. 10]) SSE: 1103.541
	•	•	•	L
K = 4	R = 3	R = 4	R = 5	R =10
K = 4 SSE	R = 3 1100.42294394	R = 4 1100.41668044	R = 5 1100.41802473	R =10 1100.42087124

K = 5	R = 3	R = 4	R = 5	R =10
SSE	773.23923763	773.19176363	773.19889256	773.19889256
K-Means	K.Means Clustering (It S. r. 3) SSE: 773.299	K. Means Clustering (b. S. r. d.) (SSE: 773.192	K.Means Clustering (I: S, r. S) SSE: 773.199	Kolmans Clintering (It S.): 10) SSE: 773.199
K = 5	R = 3	R = 4	R = 5	R =10
K = 5 SSE	R = 3 770.86081177	R = 4 770.85036529	R = 5 770.85048054	R =10 770.84733408

K = 6	R = 3	R = 4	R = 5	R =10
SSE	627.07079778	627.01725174	627.01540260	627.01540260
K-Means	K.Means Clustering (t. 6, r. 3) 55E: 677.071	K-Means Clustering (b: 6, r. 4) SSE: 627 017	K.Means Clustering (b: 6, r: 5) SSE: 627 015	K.Maans Clintering (b: 6, r: 10) SSE: 627 015
	•			
K = 6	R = 3	R = 4	R = 5	R =10
K = 6 SSE	R = 3 624.63242723	R = 4 624.62717547	R = 5 624.61702731	R =10 624.60541168

K = 7	R = 3	R = 4	R = 5	R =10
SSE	554.99363007	561.41611684	555.36501060	550.83158884
K-Means	K-Means Clustering (it-7, r. 3) (SSE: SS4 994	K.Means Clustering (k· 7, r· 4) SSE-961.416	K-Means Clustering (b: 7, r: 5) SSE: 555:365	K-Means Clustering Rr 7, r 10 SSE: 50 832
	•	•		
K = 7	R = 3	R = 4	R = 5	R =10
K = 7 SSE	R = 3 548.53399733	R = 4 551.15620980	R = 5 551.64157382	R =10 549.53175658

K = 8	R = 3	R = 4	R = 5	R =10
SSE	486.73435330	480.67830257	480.24582736	488.33771797
K-Means	K-Means Clustering (i: 8, r. 3) 55E-466,734	K-Means Clustering (it. 8, r. 4) 555: 400 678	K. Means Clustering (b: 8, r: 5) SSE: 480.246	Means Clustering (h: 8, r: 10) SSE: 488.338
K = 8	R = 3	R = 4	R = 5	R =10
K = 8 SSE	R = 3 486.13294678	R = 4 476.85155467	R = 5 477.44916000	R =10 477.20912005

K = 9	R = 3	R = 4	R = 5	R =10
SSE	433.08031480	425.61738040	418.96145999	419.30241581
K-Means	K-Means Clustering 1: 9, r. 3) [SSE: 433 (80)	K-Means Clustering (it: 9, r. 4) 555: 425-617	K-Means Clustering 1: 9, r. 5) SSE: 418 961	K-Means Clustering (ic 9 ; r 10) 55E-419-302
K = 9	R = 3	R = 4	R = 5	R =10
K = 9 SSE	R = 3 416.18131815	R = 4 416.94143682	R = 5 422.96165542	R =10 416.22197737