Cera Oh
ML
Winter 2022
3/07/2022

# Programming 3 Report

**Assignment #1: K-Means**

**Introduction**

For this assignment, I implemented K-Means algorithm as given in the lecture. The data set was provided by Dr. Rhodes. The data set contained 1500 2-dimensional points simulated from 3 Gaussians that overlap. The algorithm takes as input *k* as the number of means (also called centroid) to find. The program asks for input *r* for how many times to run the algorithm. For each run, *k* number of points are randomly chosen as initial starting means. Using the means, the algorithm calculates Euclidean distance between each point and each cluster to determine which cluster each point belongs to in the assignment() function. Then, means are recalculated using the update formula given in lecture in the recompute() function. The algorithm runs until the means no longer change. The program calculates the Sum of Squares Error (SSE) for each run, and outputs the means with minimum SSE.

Each cluster was assigned a random color by the program for plotting. Means are marked as black points. I ran the algorithm 10 times per k-size. Initial mean points, final mean points, and SSE per run were recorded.

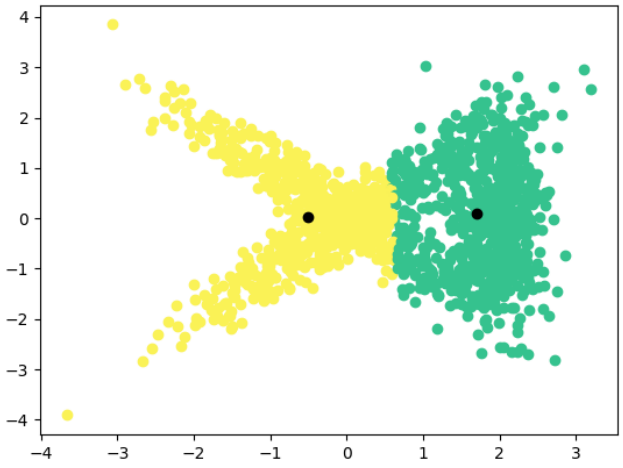
**Results**

The run with green highlight has the lowest SSE for the K. SSE was rounded to 2nd decimal place.

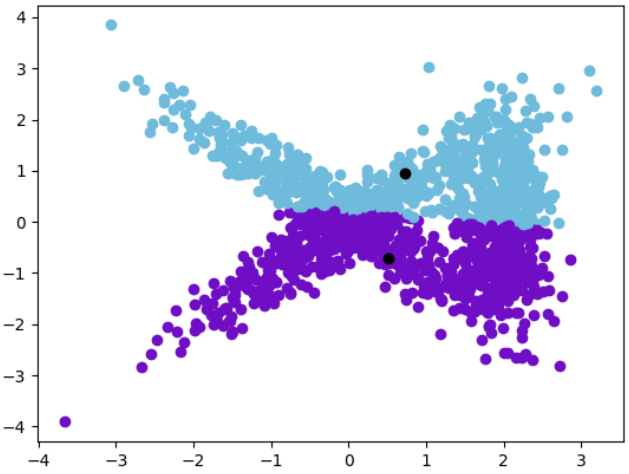
| **K = 2** | | | |
|---|---|---|---|
| **Runs** | **Initial Means** | **Final Means** | **SSE** |
| **1** | (-1.127723, 1.032879)<br>(2.328313, 1.311564) | (-0.50568675, 0.0213408)<br>(1.70232498, 0.08960845) | 2168.32 |
| **2** | (2.022325, 1.442141)<br>(-0.508743, -0.038128) | (1.70232498, 0.08960845)<br>(-0.50568675, 0.0213408) | 2168.32 |
| **3** | (2.137516, 0.404854)<br>(1.847744, 0.506489) | (1.70232498, 0.08960845)<br>(-0.50568675, 0.0213408) | 2168.32 |
| **4** | (-0.896318, 1.030859)<br>( 1.840547, -2.166207) | (-0.50568675, 0.0213408)<br>( 1.70232498, 0.08960845) | 2168.32 |
| **5** | (-0.666205, -0.535584)<br>( 2.173411, 1.233727) | (-0.50568675, 0.0213408)<br>( 1.70232498, 0.08960845) | 2168.32 |
| **6** | ( 1.822414, 1.081108) | ( 0.73227206, 0.94797395) | 2811.60 |

| | | | |
|---|---|---|---|
| | ( 1.811857, -0.831418) | ( 0.5144653, -0.70183868) | |
| **7** | ( 2.094511, 0.68652)<br>( 0.625472, -0.653344) | ( 1.70232498, 0.08960845)<br>(-0.50568675, 0.0213408) | 2168.32 |
| **8** | (-0.953005, -0.462903)<br>(-0.547845, 1.239886) | (-0.50568675, 0.0213408)<br>( 1.70232498, 0.08960845) | 2168.32 |
| **9** | (-1.256668, 1.633277)<br>( 0.625991, 0.689018) | (-0.51774696, 0.02182924)<br>( 1.69081902, 0.08843396) | 2171.04 |
| **10** | ( 2.020854, 1.225413)<br>( 1.290172, -1.00952) | ( 1.70232498, 0.08960845)<br>(-0.50568675, 0.0213408) | 2168.32 |

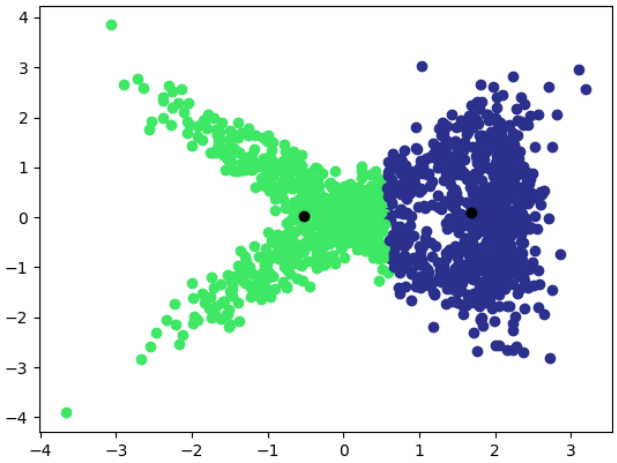
**K = 2 Run 1**



**K = 2 Run 6**



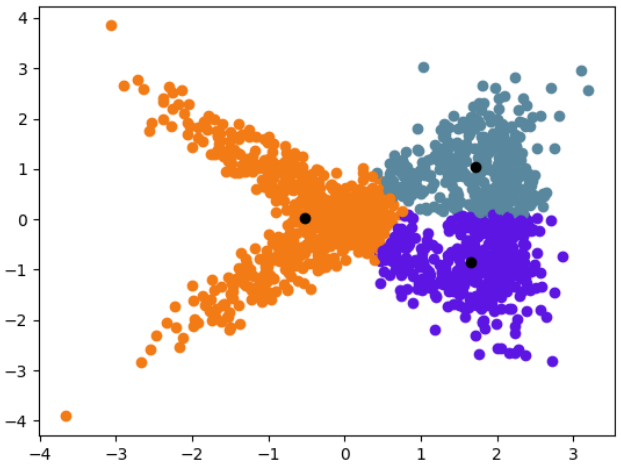
**K = 2 Run 9**



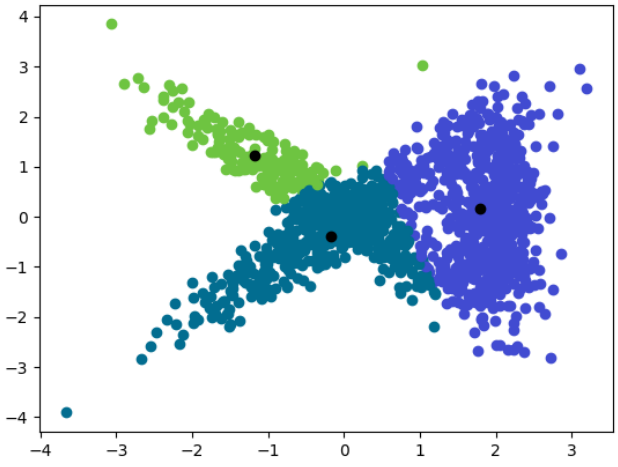
Cera Oh
ML
Winter 2022
3/07/2022

| K = 3 | | | |
|---|---|---|---|
| **Runs** | **Initial Means** | **Final Means** | **SSE** |
| **1** | (1.758259, -1.549447)<br>(-0.559493, 1.046941)<br>( 1.182248, -1.283617) | ( 1.64650399, -0.85083626)<br>(-0.52458245, 0.03628108)<br>( 1.71541534, 1.04614047) | 1478.37 |
| **2** | ( 0.879487, 0.027984)<br>(-0.439028, 0.81926 )<br>( 2.269834, -0.277637) | ( 1.71541534, 1.04614047)<br>(-0.52458245, 0.03628108)<br>( 1.64650399, -0.85083626) | 1478.37 |
| **3** | (-0.380926, -0.021427)<br>(-0.897764, 0.461577)<br>( 1.790378, -1.066175) | (-0.16822732, -0.39443595)<br>(-1.17925695, 1.23781304)<br>( 1.79402114, 0.17475807) | 1915.94 |
| **4** | (1.938256, 1.548529)<br>(2.255128, 1.27944 )<br>(1.267096, 1.323862) | ( 1.724763, 1.04927424)<br>( 1.64650399, -0.85083626)<br>(-0.52016677, 0.03882836) | 1479.54 |
| **5** | (1.938256, 1.548529)<br>(2.255128, 1.27944 )<br>(1.267096, 1.323862) | ( 1.724763, 1.04927424)<br>( 1.64650399, -0.85083626)<br>(-0.52016677, 0.03882836) | 1478.37 |
| **6** | ( 0.890813, -1.071393)<br>( 1.077669, -1.132181)<br>( 0.447464, -0.600881) | (-0.16939125, -0.3961673 )<br>( 1.79224925, 0.17553264)<br>(-1.17925695, 1.23781304) | 1913.01 |
| **7** | ( 0.810627, -0.260099)<br>( 2.648257, 0.532207)<br>( 1.764673, 1.398841) | (-0.52016677, 0.03882836)<br>( 1.64650399, -0.85083626)<br>( 1.724763, 1.04927424) | 1479.54 |
| **8** | ( 0.745527, 1.077268)<br>( 2.055097, 1.629717)<br>( 1.57403, -0.603073) | (-0.52016677, 0.03882836)<br>( 1.724763, 1.04927424)<br>( 1.64650399, -0.85083626) | 1479.54 |
| **9** | (-2.221549, -1.736306)<br>( 0.608013, 0.409494)<br>( 1.975687, -0.591943) | (-1.03628358, -1.15076271)<br>(-0.28450025, 0.40417398)<br>( 1.75003436, 0.05100056) | 1603.77 |
| **10** | (-0.507652, 0.872013)<br>( 2.234388, -0.819444)<br>(-0.321386, -0.020261) | (-1.17925695, 1.23781304)<br>( 1.79224925, 0.17553264)<br>(-0.16939125, -0.3961673 ) | 1913.01 |

**K = 3 Run 1**

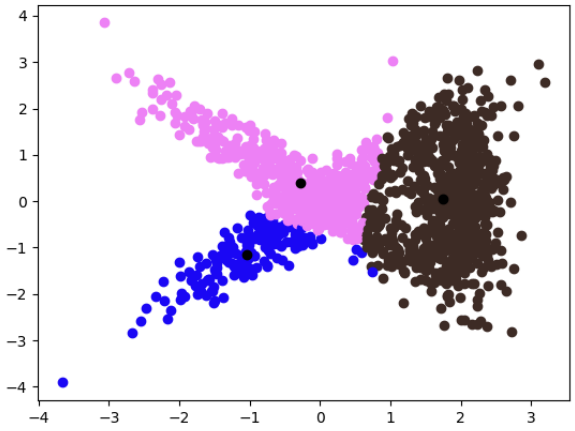


**K = 3 Run 4**



Cera Oh
ML
Winter 2022
3/07/2022

**K = 3 Run 9**

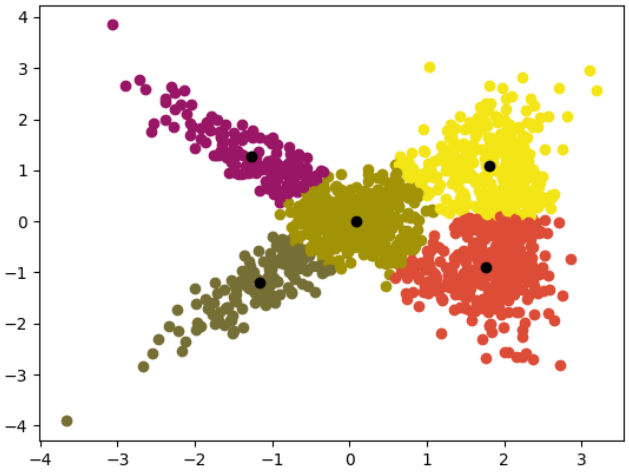


| K = 5 | | | |
|---|---|---|---|
| **Runs** | **Initial Means** | **Final Means** | **SSE** |
| **1** | (-0.951859, 1.158815)<br>( 0.043906, -0.127977)<br>( 1.536723, -1.376795)<br>(-0.243791, -0.00392 )<br>( 2.201263, 0.458657) | (-1.26993999, 1.28443945)<br>( 0.08694302, 0.00449109)<br>( 1.76389627, -0.89390295)<br>(-1.15793897, -1.20480123)<br>( 1.79890219, 1.08099452) | 781.29 |
| **2** | ( 0.427788, 0.915815)<br>( 1.69561, -0.72381 )<br>( 1.469619, 0.552333)<br>( 2.812883, 2.062197)<br>( 2.459363, 0.765557) | (-1.23309412, 0.19996993)<br>( 1.66153738, -1.20779844)<br>( 0.07621522, -0.07583429)<br>( 1.70294339, 1.41824316)<br>( 1.96958296, 0.12718118) | 1147.63 |
| **3** | ( 2.040208, 0.444612)<br>( 1.52024, 0.996018)<br>( 0.023465, -0.406911)<br>( 1.826834, 0.019837)<br>( 1.357365, -1.200832) | ( 1.96958296, 0.12718118)<br>( 1.70294339, 1.41824316)<br>(-1.23309412, 0.19996993)<br>( 0.07621522, -0.07583429)<br>( 1.66153738, -1.20779844) | 1147.63 |
| **4** | (-0.321386, -0.020261)<br>( 0.270286, 0.411257)<br>( 3.103028, 2.961063)<br>( 2.094789, -0.413662)<br>( 2.086736, -0.704343) | (-1.23309412, 0.19996993)<br>( 0.07621522, -0.07583429)<br>( 1.70294339, 1.41824316)<br>( 1.96958296, 0.12718118)<br>( 1.66153738, -1.20779844) | 1147.63 |
| **5** | ( 1.792224, 2.262982)<br>( 0.77176, 1.117794)<br>( 2.122204, -1.436975)<br>( 0.022037, 0.303651)<br>( 1.182248, -1.283617) | ( 1.70294339, 1.41824316)<br>( 1.96958296, 0.12718118)<br>( 1.66153738, -1.20779844)<br>(-1.23309412, 0.19996993)<br>( 0.07621522, -0.07583429) | 1147.63 |
| **6** | ( 1.006358, 0.511154)<br>( 1.128922, 1.204373)<br>(-0.800029, 0.719299)<br>( 2.091493, 2.453773)<br>( 0.112569, -0.304399) | ( 1.64848366, -1.19630379)<br>( 1.96510772, 0.1180974 )<br>(-1.2156474, 1.24551518)<br>( 1.58065262, 1.33029498)<br>(-0.2623308, -0.34522904) | 1289.89 |
| **7** | (-0.987872, -0.30727 )<br>( 1.808938, -2.049572)<br>( 2.328313, 1.311564)<br>( 1.452632, -1.140956) | (-1.23309412, 0.19996993)<br>( 1.66153738, -1.20779844)<br>( 1.70294339, 1.41824316)<br>( 0.07621522, -0.07583429) | 1147.63 |

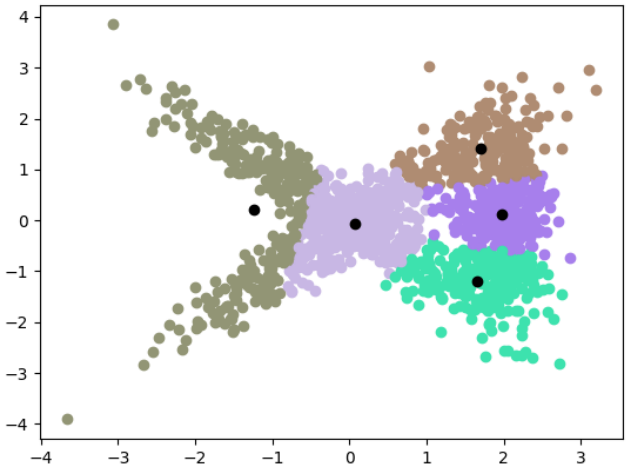
Cera Oh
ML
Winter 2022
3/07/2022

| | | | |
|---|---|---|---|
| | ( 2.138305, -0.122679) | ( 1.96958296, 0.12718118) | |
| **8** | ( 0.664539, 0.588856)<br>( 1.165228, 0.726488)<br>(-1.42041, -1.166614)<br>( 0.054442, 0.039398)<br>( 0.607328, 0.49271 ) | ( 1.69539129, 1.4120004 )<br>( 1.92657027, 0.13075258)<br>(-1.08453109, -1.15569143)<br>(-0.32635652, 0.39912141)<br>( 1.52604698, -1.12392479) | 638.67 |
| **9** | (1.883659, 1.511445)<br>(2.069274, 1.889428)<br>(1.591684, 1.166497)<br>(0.316162, 0.246654)<br>(0.360109, 0.628475) | ( 1.96510772, 0.1180974 )<br>( 1.58065262, 1.33029498)<br>( 1.64848366, -1.19630379)<br>(-0.2623308, -0.34522904)<br>(-1.2156474, 1.24551518) | 1289.89 |
| **10** | ( 1.181673, -1.492543)<br>( 2.515471, -0.672047)<br>(-0.633251, -0.048245)<br>( 2.022325, 1.442141)<br>( 1.267313, 1.201958) | ( 1.66153738, -1.20779844)<br>( 1.96958296, 0.12718118)<br>(-1.23309412, 0.19996993)<br>( 1.70294339, 1.41824316)<br>( 0.07621522, -0.07583429) | 1147.63 |

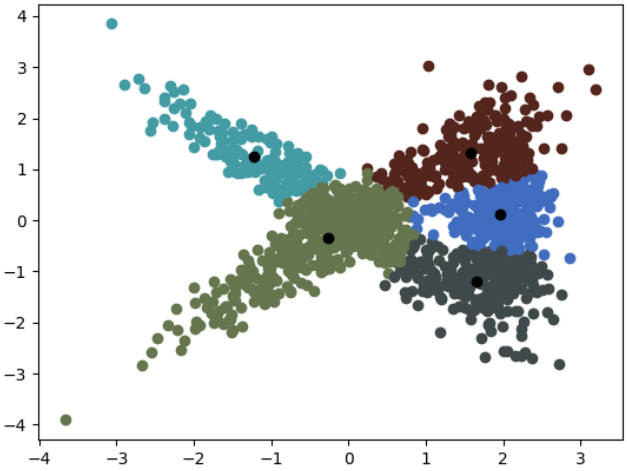
**K = 5 Run 1**



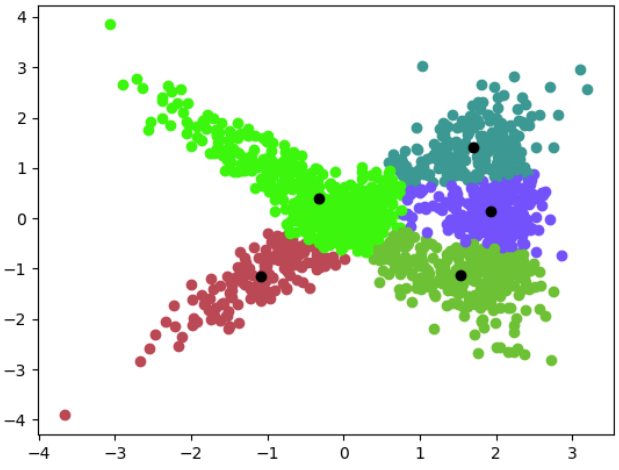
**K = 5 Run 2**



**K = 5 Run 6**



**K = 5 Run 8**

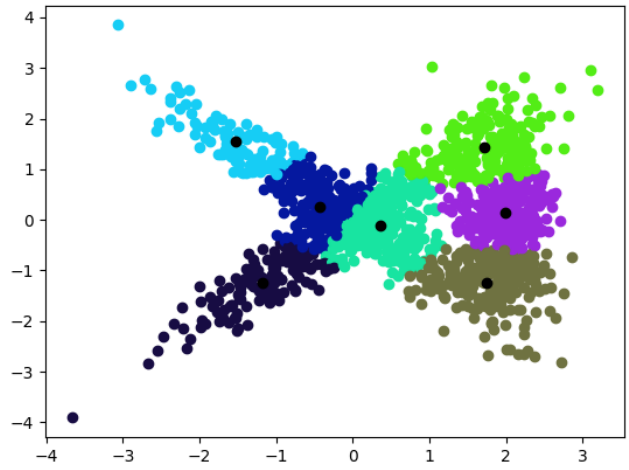


| K = 7 | | | |
|---|---|---|---|
| **Runs** | **Initial Means** | **Final Means** | **SSE** |
| **1** | ( 0.28375, 0.157829)<br>( 2.080283, 1.001673)<br>( 0.051551, -0.225876)<br>(-0.35305, -0.166191)<br>( 1.884208, -1.710103)<br>( 0.573193, 0.708433)<br>(-2.160543, -2.537075) | (-0.43050143, 0.26836883)<br>( 1.71428359, 1.44050161)<br>( 0.36158867, -0.10604409)<br>(-1.52836203, 1.53961299)<br>( 1.74356106, -1.2360507 )<br>( 1.99152793, 0.1398425 )<br>(-1.18227431, -1.25439141) | 615.08 |
| **2** | (-2.383168, 2.331112)<br>(-1.066095, -1.122648)<br>( 1.82578, 0.445319)<br>(-0.371515, 0.46515 )<br>(-1.091898, -1.631381)<br>( 1.765399, 1.9175 )<br>(-0.953005, -0.462903) | (-1.41593517, 1.42149373)<br>(-0.49755421, -0.46194774)<br>( 1.86114531, -0.89884458)<br>( 0.03099202, 0.43217113)<br>(-1.49241657, -1.56563807)<br>( 1.85545349, 1.09633262)<br>( 0.50455254, -0.39004978) | 626.88 |
| **3** | (-0.213436, 0.160775)<br>( 0.078998, 0.23628 )<br>(-0.933126, -1.18226 )<br>( 2.036062, -0.364298)<br>(-0.657486, -0.776915)<br>( 0.129257, -0.400282)<br>(-0.038093, 0.507117) | ( 0.30597727, 0.1200699 )<br>( 1.7604329, 1.45240962)<br>(-1.48021866, -1.5407465 )<br>( 1.99949393, 0.11687856)<br>(-0.45268318, -0.35911153)<br>( 1.65654618, -1.21256827)<br>(-1.25468419, 1.27356723) | 500.68 |
| **4** | ( 2.328313, 1.311564)<br>(-0.243791, -0.00392 )<br>(-1.576246, -1.681196)<br>( 2.064616, 1.713611)<br>( 0.298869, -0.069559)<br>(-0.679909, 0.577444)<br>( 1.808938, -2.049572) | ( 2.00321767, 0.11624481)<br>(-0.3620698, -0.12887468)<br>(-1.39267652, -1.45733186)<br>( 1.76937308, 1.45603817)<br>( 0.43826374, 0.10371237)<br>(-1.32968686, 1.35420147)<br>( 1.66761537, -1.21778144) | 526.89 |
| **5** | ( 1.954494, 0.248193)<br>( 1.014955, -1.413011)<br>( 2.017216, 0.291811)<br>( 1.991355, 1.480758)<br>( 0.066545, -0.335599)<br>( 1.280648, -1.023197)<br>(-1.512506, 1.545157) | (-0.00348571, 0.11022717)<br>( 0.84698177, -0.77520092)<br>( 1.96001829, 0.15670068)<br>( 1.71636919, 1.45123135)<br>(-1.15216506, -1.20238744)<br>( 1.93216915, -1.29923619)<br>(-1.35290791, 1.36043129) | 618.25 |
| **6** | (-1.46679, -1.429536)<br>(-0.484636, -0.8832 )<br>( 0.051551, -0.225876)<br>( 0.4577, -0.690819)<br>( 1.181673, -1.492543)<br>( 2.355299, -0.967471)<br>(-0.679272, -0.897629) | (-1.48568587, -1.54783361)<br>(-0.7599572, 0.79238337)<br>(-1.81786216, 1.80766402)<br>( 0.34335726, 0.06128429)<br>( 1.77829436, -0.89227915)<br>( 1.83780807, 1.10261297)<br>(-0.44641157, -0.44198221) | 643.87 |
| **7** | ( 2.158459, 1.831032)<br>( 1.417572, 0.650471)<br>( 0.086548, 0.284879)<br>(-0.646451, -0.70281 )<br>( 1.762724, 1.238707)<br>( 2.312057, -0.017315)<br>( 0.643963, -0.42177 ) | ( 1.82238514, 1.49923746)<br>( 0.56339663, 0.45182204)<br>(-1.27228733, 1.29022571)<br>(-1.30533952, -1.37027972)<br>( 2.00433491, 0.1322545 )<br>( 1.66167551, -1.20493598)<br>(-0.11119092, -0.16884632) | 489.59 |
| **8** | ( 2.182986, -0.230095)<br>( 0.187777, 0.018864)<br>( 1.847744, 0.506489)<br>(-0.666891, -0.615487)<br>( 2.245683, 1.463513) | ( 1.65884232, -1.21002746)<br>( 0.00467652, -0.07139433)<br>( 0.96632576, 0.80752498)<br>(-1.21007817, -1.26394185)<br>( 1.94094641, 1.58236024) | 501.71 |

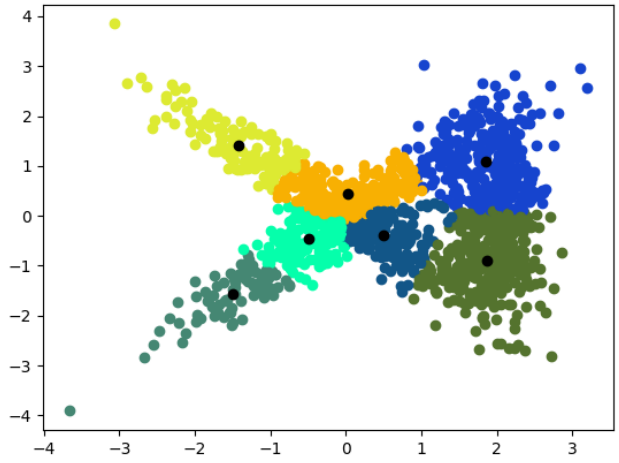
Cera Oh
ML
Winter 2022
3/07/2022

| | | | |
|---|---|---|---|
| | ( 2.018552, 0.02213 )<br>(-0.153239, 0.078505) | ( 2.01765001, 0.1128499 )<br>(-1.26993999, 1.28443945) | |
| **9** | ( 2.028785, 0.17989 )<br>( 1.63006, -1.758404)<br>( 1.665546, -1.106849)<br>( 1.841216, 0.427529)<br>( 0.397533, 0.270256)<br>(-0.433158, -0.397562)<br>(-0.469016, -0.545125) | ( 1.96166935, 0.13357317)<br>( 1.9312699, -1.30342835)<br>( 0.84698177, -0.77520092)<br>( 1.71934422, 1.42805273)<br>(-0.00558965, 0.10898447)<br>(-1.35290791, 1.36043129)<br>(-1.15216506, -1.20238744) | 616.47 |
| **10** | ( 0.54113, 0.565313)<br>( 0.127149, 0.018039)<br>( 1.865124, -0.713334)<br>( 2.42669, 1.843263)<br>( 1.727669, -1.709675)<br>(-0.214149, -0.144841)<br>( 1.765569, 0.163269) | (-0.00558965, 0.10898447)<br>(-1.15216506, -1.20238744)<br>( 0.84698177, -0.77520092)<br>( 1.71934422, 1.42805273)<br>( 1.9312699, -1.30342835)<br>(-1.35290791, 1.36043129)<br>( 1.96166935, 0.13357317) | 616.47 |

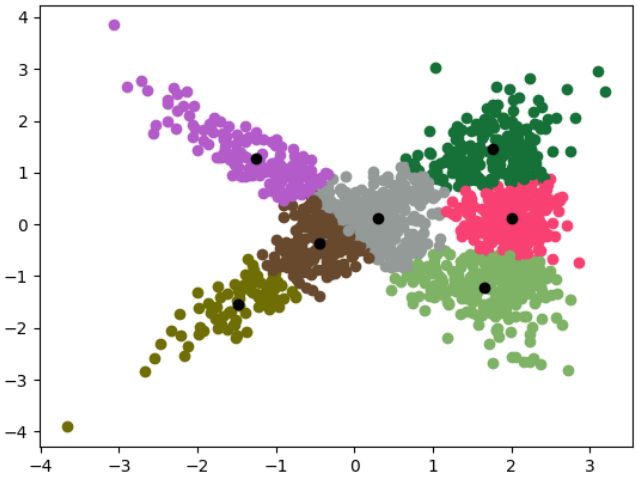
**K = 7 Run 1**



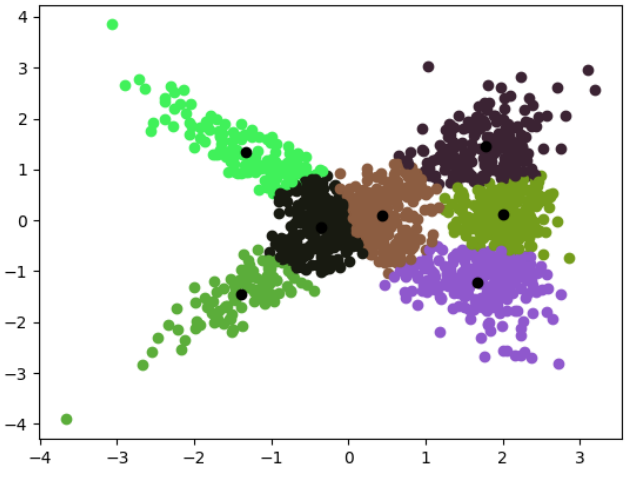
**K = 7 Run 2**



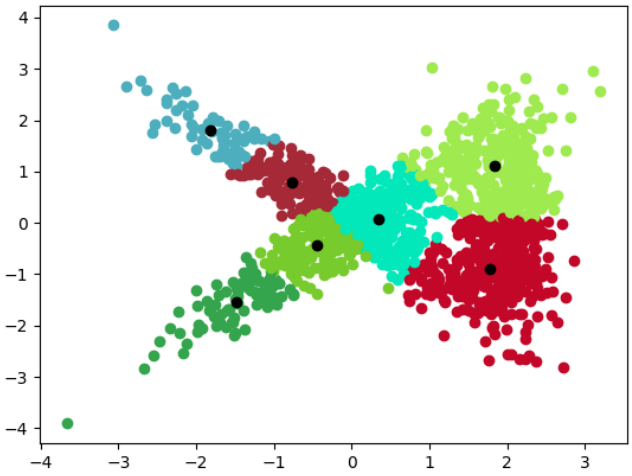
**K = 7 Run 3**



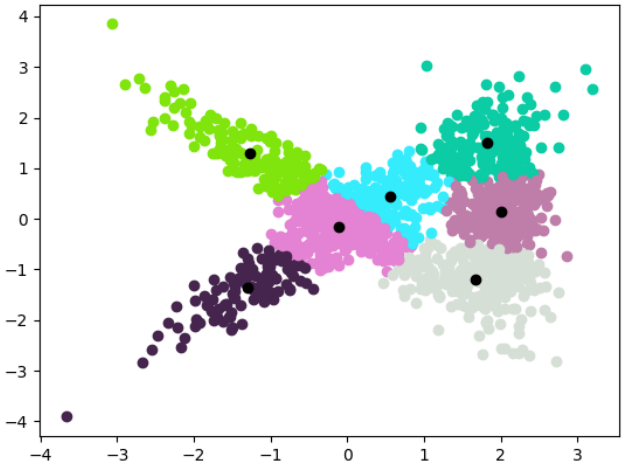
**K = 7 Run 4**



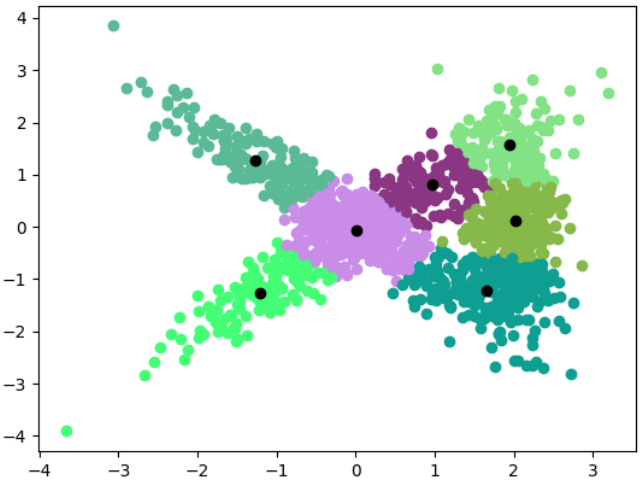
**K = 7 Run 6**



**K = 7 Run 7**



**K = 7 Run 8**

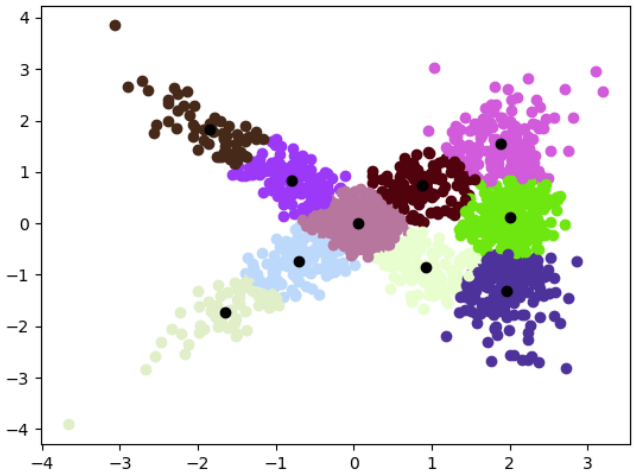


| K = 10 | | | |
|---|---|---|---|
| **Runs** | **Initial Means** | **Final Means** | **SSE** |
| **1** | (-1.066095, -1.122648)<br>(-1.600471, 1.906975)<br>( 0.21732, -0.528494)<br>( 2.277366, 2.164012)<br>(-0.274246, 0.081864)<br>(-0.708482, -0.397262)<br>( 0.029899, -0.263089)<br>( 0.308016, 0.244693)<br>( 0.25935, -0.550914) | (-1.64687839, -1.71840192)<br>(-1.84995692, 1.82564092)<br>( 0.93060464, -0.84685457)<br>( 1.88472012, 1.53953961)<br>(-0.79697964, 0.83036007)<br>(-0.70668514, -0.73792076)<br>( 0.04949386, -0.0070501 )<br>( 0.87569649, 0.73605297)<br>( 1.96017091, -1.30677434) | 372.72 |

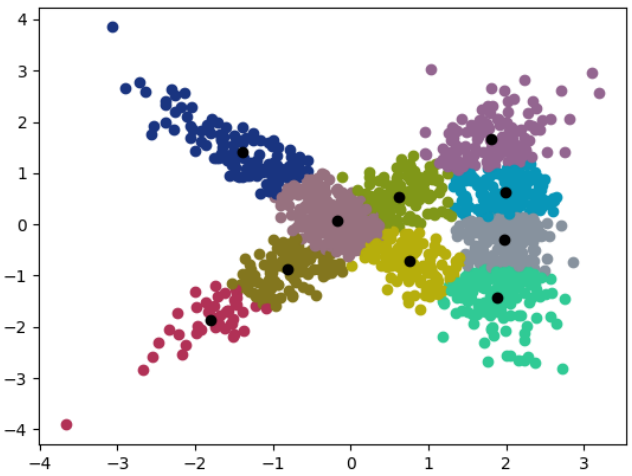
| | | | |
|---|---|---|---|
| | ( 0.306134, -0.152084) | ( 2.00981534, 0.12402325) | |
| **2** | ( 2.109455, 1.316184)<br>( 1.182248, -1.283617)<br>(-2.122847, -2.34787 )<br>( 1.806806, -2.021413)<br>( 2.017216, 0.291811)<br>( 2.415148, 0.899712)<br>(-0.910405, -1.429968)<br>(-0.233936, -0.250617)<br>(-0.667771, -0.689108)<br>(-1.057451, 0.642733) | ( 1.80004531, 1.66486774)<br>( 0.74981314, -0.7218353 )<br>(-1.8052775, -1.87209357)<br>( 1.8759982, -1.43977544)<br>( 1.97994352, -0.30725888)<br>( 1.98777859, 0.62410376)<br>(-0.8166554, -0.87136098)<br>( 0.61380832, 0.52668551)<br>(-0.17953682, 0.07818272)<br>(-1.39541511, 1.4139271 ) | 333.43 |
| **3** | ( 0.831881, -0.280967)<br>( 2.084048, 1.053426)<br>( 0.023465, -0.406911)<br>( 0.201314, -0.080247)<br>(-0.962856, 0.757454)<br>( 0.185216, 0.294869)<br>( 1.694928, 0.485462)<br>( 0.823293, 1.04636 )<br>( 1.267096, 1.323862)<br>( 0.940472, 0.811664) | ( 1.61655537, -1.34471182)<br>( 2.03477511, 0.60696256)<br>(-1.49241657, -1.56563807)<br>(-0.52302085, -0.47891792)<br>(-1.81786216, 1.80766402)<br>(-0.7572979, 0.78825889)<br>( 1.9752045, -0.34740749)<br>( 0.95267948, 0.86501818)<br>( 1.91276833, 1.76096477)<br>( 0.24950389, -0.07332678) | 340.36 |
| **4** | (-0.63345, 0.091873)<br>( 1.727669, -1.709675)<br>( 0.734494, -1.116504)<br>( 2.454316, 2.037399)<br>(-0.897764, 0.461577)<br>(-0.144672, 0.206945)<br>(-0.169513, -0.24397 )<br>( 1.521703, 0.346761)<br>( 0.279178, 0.004785)<br>( 1.542505, 0.064365) | (-1.49485543, -1.57530423)<br>( 1.96287619, -1.3091775 )<br>( 0.93680389, -0.85274223)<br>( 1.93256228, 1.61701778)<br>(-1.84995692, 1.82564092)<br>(-0.7903355, 0.82426187)<br>(-0.56739613, -0.54473126)<br>( 1.02306738, 0.87317172)<br>( 0.1299364, 0.03102807)<br>( 2.01751, 0.14021274) | 383.01 |
| **5** | ( 0.771715, -0.522658)<br>( 1.523657, -0.705424)<br>( 1.913897, -0.364716)<br>( 1.604107, 1.898835)<br>( 1.085916, -0.268067)<br>(-0.418261, -0.865379)<br>( 0.299749, -0.413927)<br>( 1.577473, -1.032292)<br>( 2.020069, -1.104972)<br>(-0.721023, -0.285133) | (-0.12965223, -0.1943843 )<br>( 1.97734567, -0.35985613)<br>( 1.98913594, 0.58938329)<br>( 1.80510955, 1.65847999)<br>( 0.58331029, 0.49400768)<br>(-1.30533952, -1.37027972)<br>(-0.7719675, 0.80714866)<br>( 0.9457367, -0.88319244)<br>( 1.94987699, -1.51678391)<br>(-1.81786216, 1.80766402) | 424.72 |
| **6** | ( 1.004597, -0.869738)<br>( 0.187777, 0.018864)<br>( 1.886389, -0.71555 )<br>(-1.006862, 1.177127)<br>( 2.368292, 1.021828)<br>( 0.59089, -1.116148)<br>( 1.791264, -1.405776)<br>( 0.008839, -0.038606)<br>( 1.452016, -1.088161)<br>( 0.944731, 0.812498) | ( 0.09565416, -0.00991879)<br>(-0.77014108, 0.7965132 )<br>( 2.00846134, 0.1163655 )<br>(-1.81786216, 1.80766402)<br>( 1.94094641, 1.58236024)<br>(-1.6382365, -1.71122923)<br>( 1.96691401, -1.3195062 )<br>(-0.69393927, -0.70804277)<br>( 0.96039411, -0.86243363)<br>( 0.9830197, 0.85790068) | 373.38 |
| **7** | (-0.79563, 0.370799)<br>( 1.938256, 1.548529)<br>( 1.964647, 0.315414)<br>( 0.49134, 0.256124)<br>( 0.806786, -0.366942)<br>( 0.24269, 0.919991) | (-1.49485543, -1.57530423) ( 1.90485717, 1.76141731)<br>( 1.9752045, -0.34740749)<br>( 0.2394046, -0.07484267)<br>( 1.61655537, -1.34471182)<br>(-1.81786216, 1.80766402) | 338.31 |

| | | | |
|---|---|---|---|
| | ( 0.635179, 1.122877)<br>( 2.52014, 1.413701)<br>(-0.044873, 0.401598)<br>( 0.134562, 0.215906) | ( 0.95272397, 0.8566747 )<br>( 2.03477511, 0.60696256)<br>(-0.75304035, 0.77961508)<br>(-0.54280508, -0.50462684) | |
| **8** | ( 1.826756, 2.304031)<br>(-0.691757, 0.674129)<br>( 1.716559, 0.574402)<br>( 1.737585, -0.869973)<br>( 2.233337, -2.649445)<br>( 1.886389, -0.71555 )<br>(-0.192206, -0.259271)<br>( 1.594825, -0.454304)<br>( 0.336405, -0.191949)<br>( 1.106773, 0.858656) | ( 1.86578865, 1.77277179)<br>(-1.35629555, 1.36644137)<br>( 2.07083824, 0.7654123 )<br>( 0.82871124, -0.80339963)<br>( 1.89402555, -1.73902691)<br>( 1.95347196, -0.88872411)<br>(-1.21820978, -1.28233169)<br>( 1.94761504, -0.0282297 )<br>(-0.10949372, 0.01673301)<br>( 0.89405104, 0.77178661) | 419.07 |
| **9** | ( 0.308016, 0.244693)<br>(-0.330372, 0.627528)<br>( 0.012377, -0.808539)<br>( 0.127149, 0.018039)<br>(-0.333941, -0.070463)<br>(-0.35305, -0.166191)<br>( 0.189049, -0.060512)<br>( 2.018117, 0.824276)<br>( 1.897022, -1.229894)<br>( 0.629274, 0.774093) | ( 0.77178324, 0.69403132)<br>(-1.81786216, 1.80766402)<br>(-1.67328114, -1.740469 )<br>( 0.04030979, -0.04349738)<br>(-0.7719675, 0.80714866)<br>(-0.73092043, -0.77193315)<br>( 0.95640623, -0.85948623)<br>( 2.00335562, 0.15003643)<br>( 1.96602965, -1.3105418 )<br>( 1.85386203, 1.54720801) | 376.19 |
| **10** | (-0.032747, -0.261643)<br>( 2.355299, -0.967471)<br>(-1.993166, 1.444197)<br>(-0.865994, -0.33212 )<br>( 2.018552, 0.02213 )<br>( 0.578651, -0.621566)<br>(-1.162251, 1.636383)<br>( 1.948857, 1.406676)<br>(-0.268191, -0.268986)<br>(-1.030857, 1.074553) | ( 0.42157064, 0.33935843)<br>( 1.93946678, -1.30238152)<br>(-2.25527968, 2.23653252)<br>(-1.40283356, -1.47060048)<br>( 1.98641229, 0.12906441)<br>( 0.88626707, -0.8153685 )<br>(-1.34907915, 1.3440847 )<br>( 1.80765628, 1.48195395)<br>(-0.27436169, -0.32694047)<br>(-0.62016584, 0.65860093) | 467.46 |

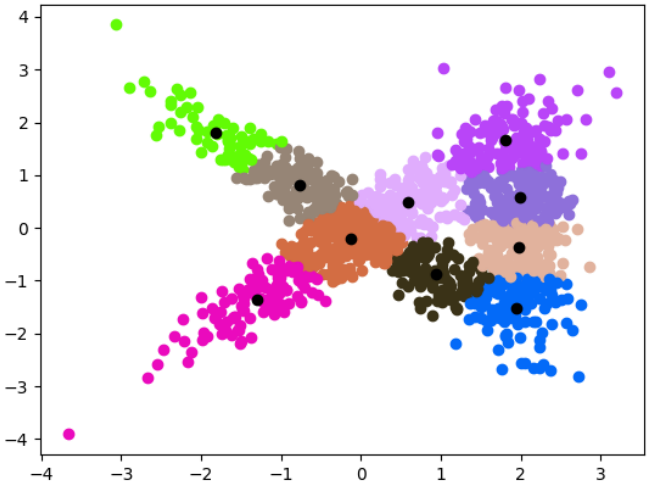
**K = 10 Run 1**



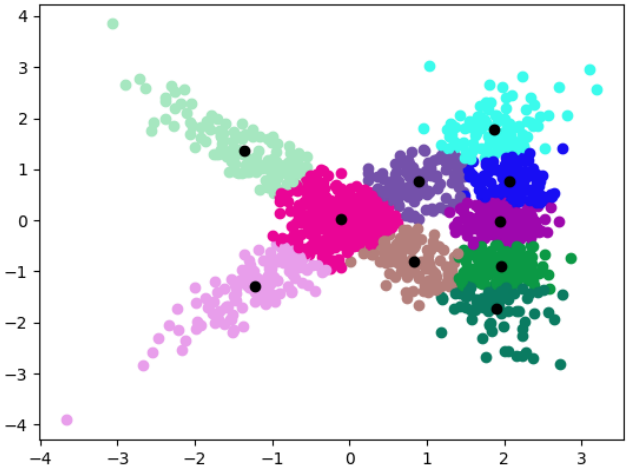

**K = 10 Run 2**



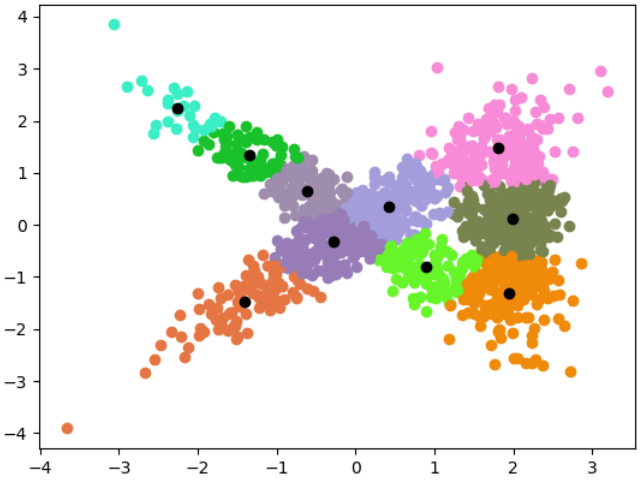
**K = 10 Run 5**



**K = 10 Run 8**



**K = 10 Run 10**



**Analysis**

For K=2, the range for SSE was 2168-2812. Lowest SSE was 2168.32. Random starting means did effect how the data points are split into two clusters, but most SSE values were around 2168.32. As the graphs show, where the initial starting means were changed the way the data is split into clusters. For example, Run 1 and Run 9 graph shows vertical splitting of data points into right and left clusters. Run 6 graph shows horizontal splitting of data points into up and below clusters. For K = 3, the range for SSE was 1478.37-1915.95. The lowest SSE was

1478.37. There was more variation in SSE values in K=3 than in K=2 as initial means had more effect on where the final means ended up. Three graphs are provided to show some of the possible clusters. K=5 had more variation in SSE and larger range of values than the previous two. SSE ranged from 638-1290. Lowest SSE was 638.67. Since we have 5 means, there were more ways to split data into 5 different clusters, as can be seen in the graphs. K=7 showed less repetition in SSE values. The SSE range was 489-644. Lowest SSE was 489.59. Initial randomly chosen mean values seems to affect the final clusters more since clusters that are formed are smaller. K=10 had SSE range of 333-468 with lowest SSE at 333.43. K=10 shows even more variation in SSE values and clusters that are formed.

**Conclusion**

SSE gets lower as the number of clusters, *k*, increases since it becomes easier to form tighter clusters as number of centroids increases. As the number of centroid increases, there seems to be also more variation in SSE values we get from different runs and SSE value seems to be affected by the initial randomly chosen means.

**Assignment #2: Fuzzy C-Means**

**Introduction**

For this assignment, I implemented Fuzzy C-Means algorithm as given in the lecture. The data set was provided by Dr. Rhodes. The data set is same as that of experiment 1. The algorithm takes as input *c* as the number of centroids to find. The program asks for input *r* for how many times to run the algorithm. Fuzzifier parameter, *m*, was set to 2. For each run, grade (also known as weight) matrix is initialized with zeros and then filled with a value between 0.0 and 1.0 by utilizing the random number generator. For each point, all grades for the clusters add up to 1.0. Using the weights, the algorithm calculates centroids using the formula given on the assignment sheet and the lecture in the centroid_finder() function. Then, weights are recalculated using the update formula given in lecture and the assignment sheet using the recompute_w() function. The algorithm runs until the centroids no longer change. The centroid was rounded up to $4^{th}$ decimal place as the cut-off point. If values between old centroids and

the newly calculated centroids are the same up to 4th decimal place, the run is complete. The program finds the Sum of Squares Error (SSE) for each run, and outputs the means with minimum SSE.

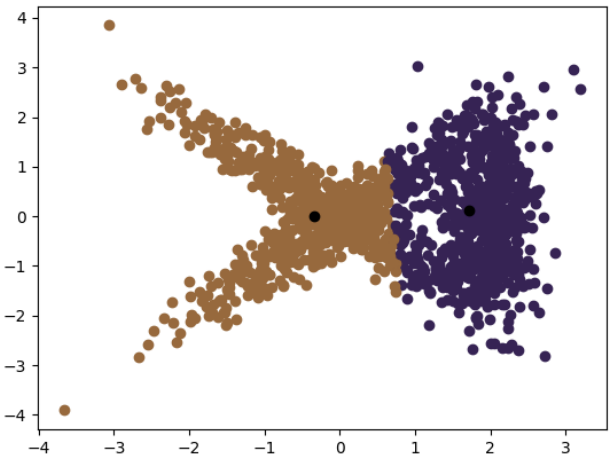
Each cluster was assigned a random color by the program for plotting. Each point was colored to reflect the cluster with highest grade for plotting. SSE was calculated using the highest-grade cluster for the point. However, many points at edges of clusters belong to more than one cluster. Centroids are marked as black points. I ran the algorithm 10 times per c-size. The final centroids and SSE were recorded per run.

**Results**

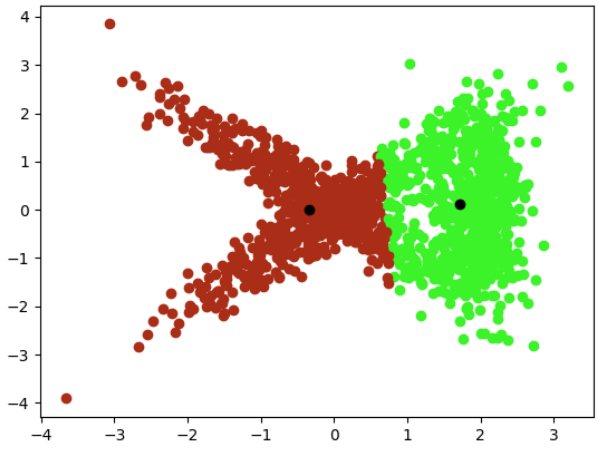
The run with green highlight has the lowest SSE for the C. SSE was rounded to 2nd decimal place.

| **C = 2** | | |
|---|---|---|
| **Runs** | **Final Centroids** | **SSE** |
| **1** | (-0.3379, 0.0036) ( 1.7088, 0.1236) | 2150.27 |
| **2** | ( 1.7088, 0.1237) (-0.3379, 0.0035) | 2150.25 |
| **3** | ( 1.7088, 0.1236) (-0.3379, 0.0036) | 2150.25 |
| **4** | ( 1.7088, 0.1237) (-0.3379, 0.0035) | 2150.25 |
| **5** | ( 1.7088, 0.1237) (-0.3379, 0.0035) | 2150.25 |
| **6** | ( 1.7088, 0.1236) (-0.3379, 0.0036) | 2150.27 |
| **7** | ( 1.7088, 0.1236) (-0.3379, 0.0036) | 2150.27 |
| **8** | ( 1.7088, 0.1236) (-0.3379, 0.0036) | 2150.27 |
| **9** | (-0.3379, 0.0035) ( 1.7088, 0.1237) | 2150.25 |
| **10** | ( 1.7088, 0.1236) (-0.3379, 0.0036) | 2150.27 |

**C = 2 Run 1**

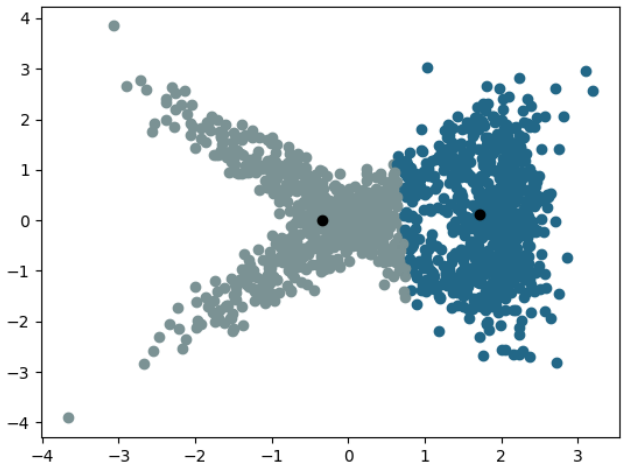



**C = 2 Run 9**



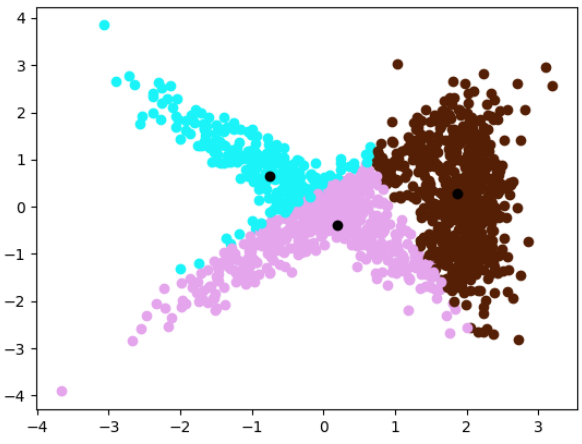
Cera Oh
ML
Winter 2022
3/07/2022

**C = 2 Run 10**

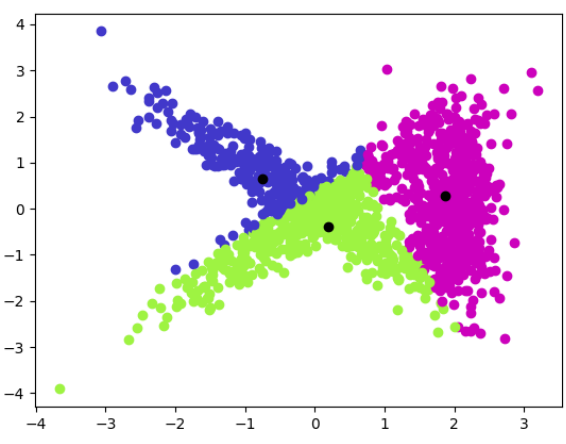


| C = 3 | | |
|---|---|---|
| **Runs** | **Final Centroids** | **SSE** |
| **1** | (-0.7455, 0.6479) (1.8677, 0.2858) (0.1957, -0.3926) | 1830.58 |
| **2** | (0.1969, -0.3932) (-0.7449, 0.6465) (1.8677, 0.2869) | 1823.78 |
| **3** | (1.6415, 1.0372) (1.5528, -0.8613) (-0.4389, 0.0257) | 1486.39 |
| **4** | ( 1.5528, -0.8613) (-0.4389, 0.0257) ( 1.6415, 1.0372) | 1486.39 |
| **5** | ( 1.6414, 1.0373) ( 1.5529, -0.8612) (-0.4389, 0.0256) | 1486.36 |
| **6** | ( 1.5529, -0.8612) (-0.4389, 0.0256) ( 1.6414, 1.0373) | 1486.36 |
| **7** | ( 1.6414, 1.0373) ( 1.5529, -0.8612) (-0.4389, 0.0256) | 1486.36 |
| **8** | ( 1.8677, 0.2869) (-0.7449, 0.6465) ( 0.1969, -0.3932) | 1823.78 |
| **9** | (1.6414, 1.0373) ( 1.5529, -0.8612) (-0.4389, 0.0256) | 1486.36 |
| **10** | (0.1954, -0.3924) (1.8677, 0.2855) (-0.7456, 0.6483) | 1829.97 |

**C = 3 Run 1**

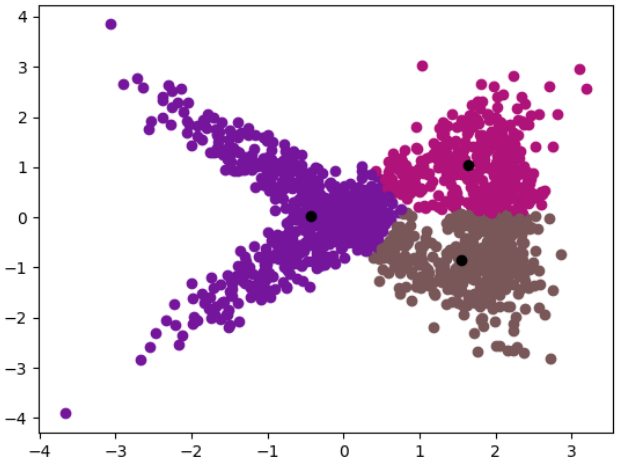


**C = 3 Run 3**

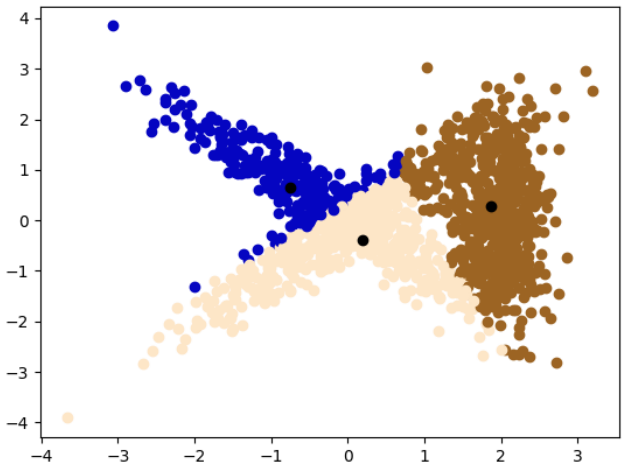


Cera Oh
ML
Winter 2022
3/07/2022

**C = 3 Run 7**



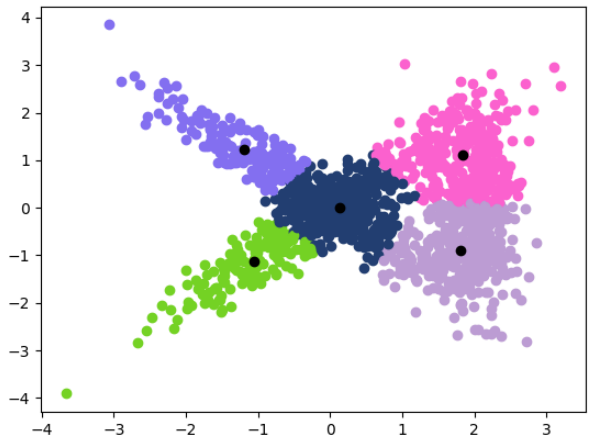
**C = 3 Run 10**



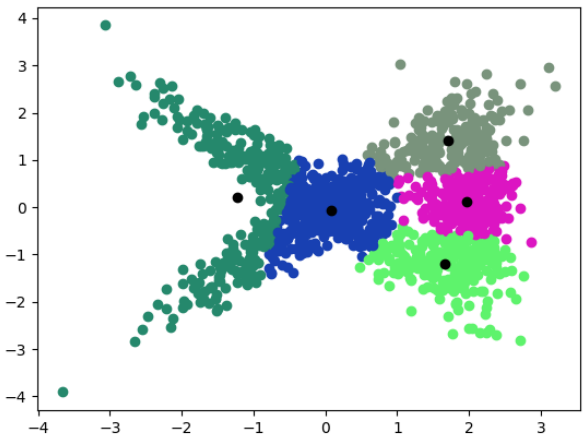
| C = 5 | | |
|---|---|---|
| **Runs** | **Final** | **SSE** |
| **1** | (-1.1974, 1.223 ) (-1.0553, -1.1214) ( 1.8395, 1.1094) ( 0.1274, -0.0043) ( 1.8137, -0.8949) | 783.02 |
| **2** | ( 1.8137, -0.8949) ( 0.1274, -0.0042) ( 1.8395, 1.1094) (-1.1974, 1.223 ) (-1.0553, -1.1214) | 783.02 |
| **3** | (-1.0553, -1.1214) ( 0.1274, -0.0043) ( 1.8137, -0.8949) (-1.1974, 1.223 ) ( 1.8395, 1.1094) | 783.02 |
| **4** | ( 0.1274, -0.0042) ( 1.8395, 1.1094) (-1.1974, 1.223 ) ( 1.8137, -0.8949) (-1.0552, -1.1214) | 783.03 |
| **5** | (-0.964, -0.8747) ( 1.6299, -1.1516) (-0.2102, 0.2798) ( 1.8544, 0.1338) ( 1.6911, 1.4422) | 669.86 |
| **6** | (-1.1973, 1.2229) (-1.0554, -1.1215) ( 1.8395, 1.1094) ( 0.1274, -0.0043) ( 1.8137, -0.8949) | 783.01 |
| **7** | (-1.0553, -1.1215) ( 0.1274, -0.0043) ( 1.8137, -0.8949) (-1.1974, 1.223) ( 1.8395, 1.1094) | 783.02 |
| **8** | ( 1.8137, -0.8949) (-1.0553, -1.1215) ( 1.8395, 1.1094) (-1.1974, 1.223) ( 0.1274, -0.0043) | 783.02 |
| **9** | ( 0.1274, -0.0043) (-1.1973, 1.2229) ( 1.8395, 1.1094) (-1.0554, -1.1216) ( 1.8137, -0.8949) | 783.01 |
| **10** | ( 1.8137, -0.8949) ( 0.1274, -0.0042) (-1.0552, -1.1214) ( 1.8395, 1.1094) (-1.1975, 1.2231) | 783.03 |

Cera Oh
ML
Winter 2022
3/07/2022

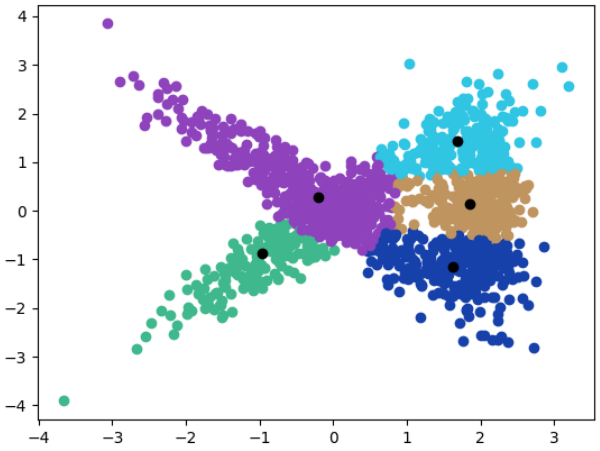
**C = 5 Run 1**



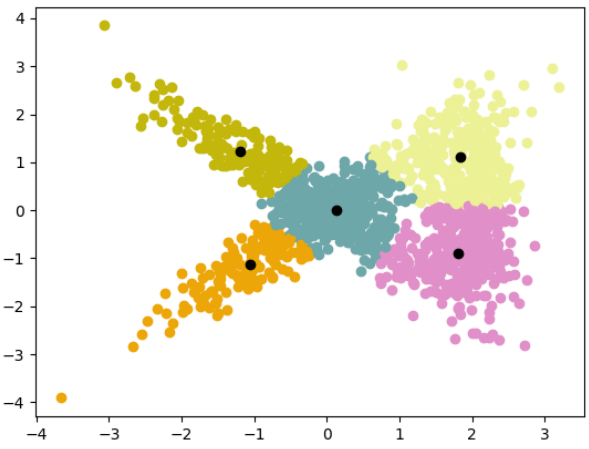
**C = 5 Run 2**



**C = 5 Run 5**



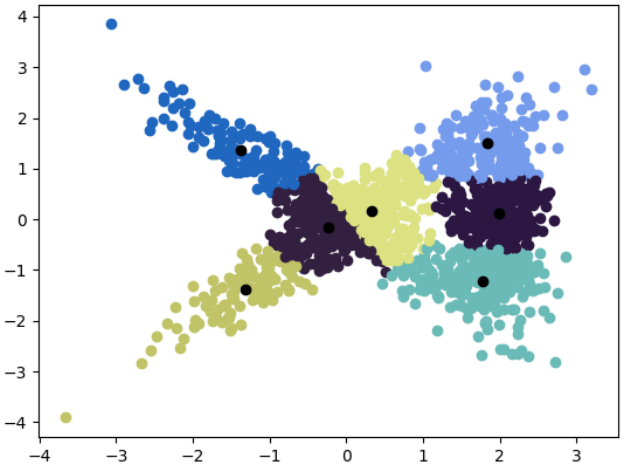
**C = 5 Run 10**



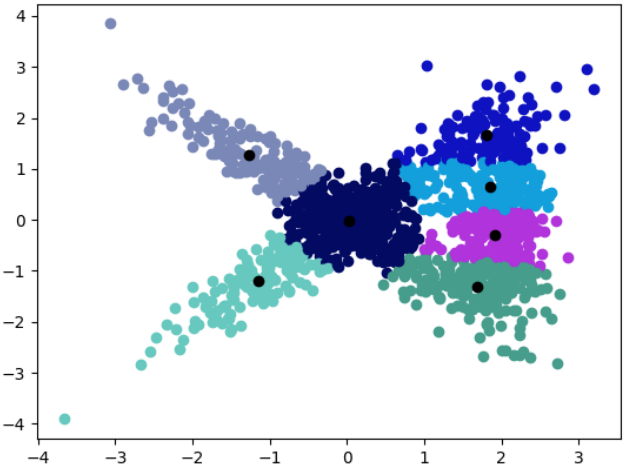
| **C = 7** | | |
|---|---|---|
| **Runs** | **Final Centroids** | **SSE** |
| **1** | ( 0.3336, 0.1613) ( 1.9814, 0.1223) (-0.2408, -0.1687) (-1.3216, -1.3831) ( 1.83, 1.4979) (-1.378, 1.3774) ( 1.775, -1.2271) | 515.91 |
| **2** | ( 0.02, -0.013 ) ( 1.687, -1.3151) ( 1.8085, 1.6656) (-1.1533, -1.1989) ( 1.8447, 0.6524) (-1.2671, 1.2672) ( 1.9102, -0.2896) | 551.93 |

| | | |
|---|---|---|
| **3** | ( 1.9758, 0.1402) ( 0.3058, -0.1766) (-1.4623, 1.4552) ( 1.8162, 1.4911) (-0.2617, 0.1946) ( 1.7954, -1.2282) (-1.2582, -1.3192) | 623.58 |
| **4** | ( 1.83, 1.4979) ( 0.3337, 0.1611) (-1.3216, -1.3831) ( 1.775, -1.2271) (-0.2409, -0.1685) ( 1.9814, 0.1224) (-1.378, 1.3774) | 515.92 |
| **5** | ( 0.3058, -0.1767) (-1.2582, -1.3192) (-1.4623, 1.4552) ( 1.8162, 1.4911) (-0.2616, 0.1946) ( 1.9758, 0.1402) ( 1.7954, -1.2282) | 623.58 |
| **6** | ( 1.775, -1.2271) ( 0.3337, 0.1611) ( 1.9814, 0.1224) (-0.2409, -0.1685) (-1.3216, -1.3831) ( 1.83, 1.4979) (-1.378, 1.3774) | 515.92 |
| **7** | (-1.4618, 1.4547) (-0.2619, 0.1932) ( 1.8162, 1.4911) ( 0.3066, -0.1759) ( 1.7954, -1.2283) ( 1.9759, 0.1402) (-1.2584, -1.3194) | 623.46 |
| **8** | ( 0.3343, 0.1594) (-0.2417, -0.1671) ( 1.9814, 0.1225) (-1.3782, 1.3776) ( 1.8299, 1.4979) (-1.3213, -1.3829) ( 1.775, -1.2271) | 516.00 |
| **9** | (-1.2584, -1.3194) ( 1.7954, -1.2283) ( 1.9759, 0.1402) (-0.2619, 0.193 ) ( 0.3067, -0.1758) (-1.4617, 1.4546) ( 1.8162, 1.4912) | 623.45 |
| **10** | (-0.2619, 0.1934) (-1.2584, -1.3194) ( 1.8162, 1.4911) ( 0.3065, -0.1759) (-1.4619, 1.4548) ( 1.9759, 0.1402) ( 1.7954, -1.2283) | 623.47 |

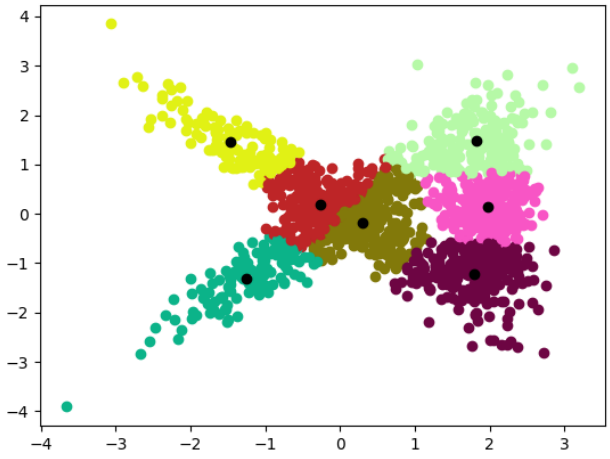
**C = 7 Run 1**



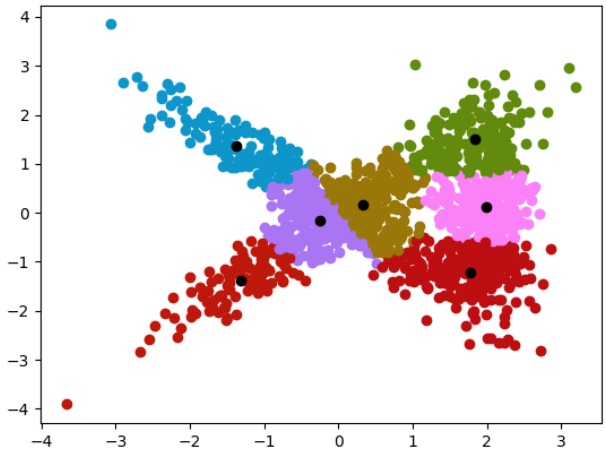

**C = 7 Run 2**



**C = 7 Run 3**



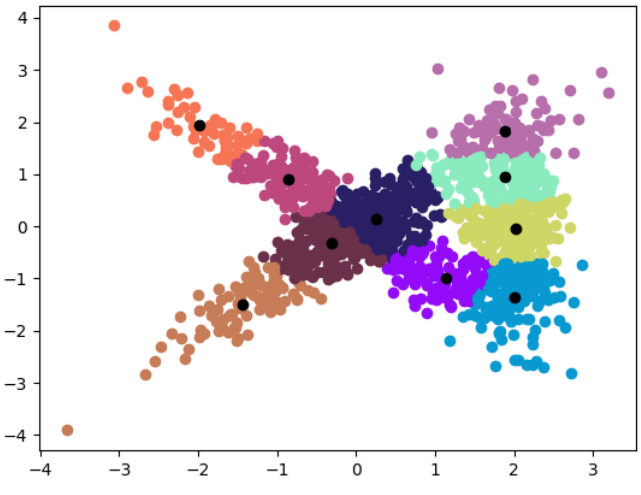
**C = 7 Run 8**



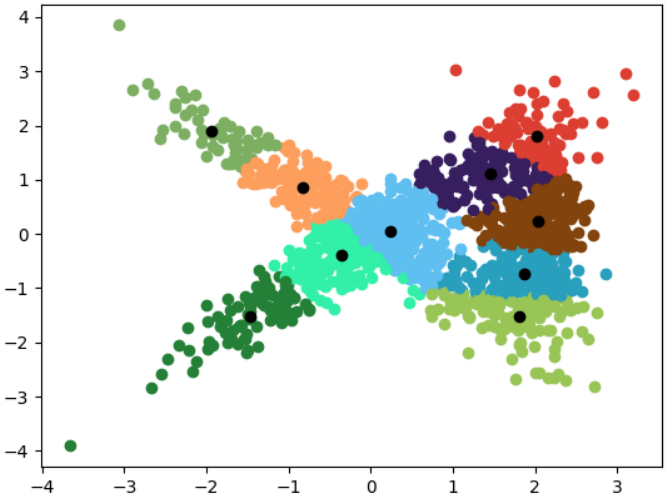
| C = 10 | | |
|---|---|---|
| **Runs** | **Final Centroids** | **SSE** |
| **1** | ( 1.1315, -0.9902) ( 2.0141, -0.0461) ( 2.01, -1.3585) (-0.3079, -0.3191) ( 1.884, 0.943 ) (-1.4427, -1.504 ) ( 1.8895, 1.8353) (-1.9768, 1.942 ) (-0.858, 0.8936) ( 0.2563, 0.1459) | 424.44 |
| **2** | (-0.8212, 0.8571) ( 0.2446, 0.0384) (-0.3552, -0.3883) (-1.9414, 1.9072) ( 1.4582, 1.1046) ( 1.803, -1.518 ) ( 2.0415, 0.2448) ( 1.8618, -0.738 ) (-1.4641, -1.5302) ( 2.0168, 1.8104) | 392.65 |
| **3** | ( 2.0141, -0.046 ) ( 2.01, -1.3585) ( 1.1315, -0.9902) ( 0.2561, 0.1457) (-1.443, -1.5043) ( 1.8896, 1.8354) (-0.3083, -0.3195) (-0.8579, 0.8935) ( 1.8838, 0.9431) (-1.9766, 1.9419) | 424.44 |
| **4** | ( 0.2563, 0.1459) (-1.4427, -1.504 ) ( 1.884, 0.943 ) ( 1.1315, -0.9902) (-0.3079, -0.3191) (-1.9768, 1.942 ) ( 1.8895, 1.8353) (-0.858, 0.8936) ( 2.01, -1.3585) ( 2.0141, -0.0461) | 424.44 |
| **5** | ( 1.1315, -0.9902) ( 0.2563, 0.1459) (-1.4427, -1.504 ) (-0.858, 0.8936) (-0.3079, -0.3191) (-1.9768, 1.942 ) ( 1.884, 0.943 ) ( 1.8895, 1.8353) ( 2.0141, -0.0461) ( 2.01, -1.3585) | 424.44 |
| **6** | ( 1.9801, -0.3252) ( 0.4521, 0.3761) (-1.9891, 1.9577) | 358.22 |

| | | |
|---|---|---|
| | (-0.8621, 0.8951) (-0.7135, -0.7494) ( 0.0612, -0.1698) (-1.697, -1.7819) ( 1.7907, -1.368 ) ( 1.8736, 1.7258) ( 2.0168, 0.6519) | |
| **7** | (-0.307, -0.3183) (-1.4421, -1.5034) (-0.8583, 0.8938) ( 0.2567, 0.1465) ( 1.1312, -0.9902) ( 2.0141, -0.0465) ( 1.8846, 0.9425) (-1.9771, 1.9423) ( 1.8893, 1.8351) ( 2.0099, -1.3587) | 424.44 |
| **8** | (-0.8583, 0.8938) (-0.307, -0.3183) ( 0.2567, 0.1465) ( 1.8846, 0.9425) (-1.4421, -1.5034) ( 2.0141, -0.0465) ( 1.8893, 1.8351) ( 1.1312, -0.9902) ( 2.0099, -1.3587) (-1.9771, 1.9423) | 424.44 |
| **9** | (-1.9769, 1.9421) (-0.8581, 0.8937) (-0.3075, -0.3188) (-1.4425, -1.5038) ( 1.1315, -0.9903) ( 2.0141, -0.0461) ( 0.2565, 0.1461) ( 2.01, -1.3585) ( 1.8895, 1.8354) ( 1.8841, 0.943 ) | 424.45 |
| **10** | ( 0.2455, 0.0392) ( 2.0414, 0.2434) (-0.3528, -0.3859) (-1.4626, -1.5286) (-0.8218, 0.8576) ( 2.0167, 1.8111) ( 1.8605, -0.7395) ( 1.4606, 1.1046) ( 1.804, -1.519 ) (-1.9419, 1.9077) | 392.79 |

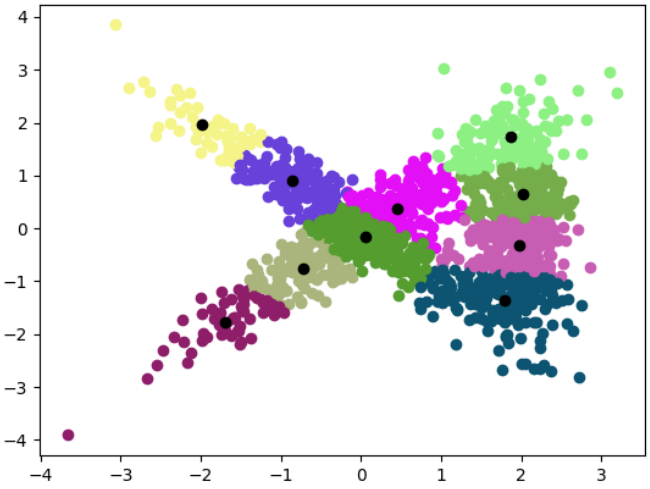
**C = 10 Run 1**



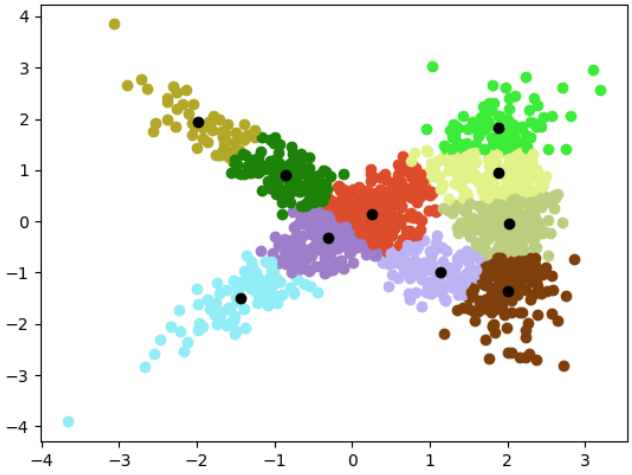

**C = 10 Run 2**



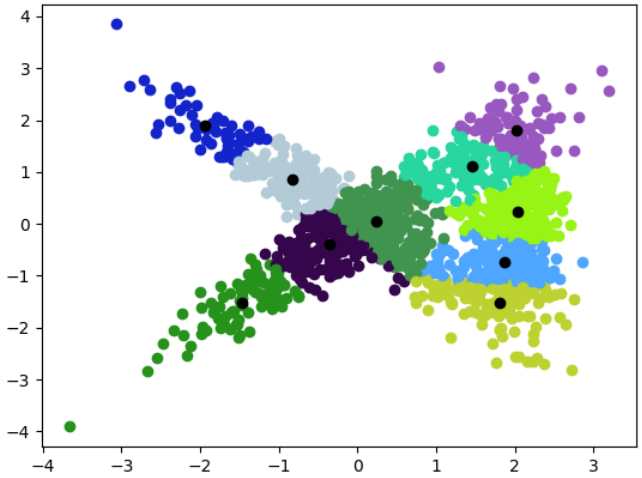
**C = 10 Run 6**



**C = 10 Run 9**



**C = 10 Run 10**



**Analysis**

For C=2, the range for SSE was 2150.25-2150.27. This is a smaller range than what we saw for K-Means, which was 2168-2812. Lowest SSE was 2150.25. Since Fuzzy means is more flexible and not as sensitive to outliers, it was able to categorize points to one group or more by grades and therefore had smaller SSE. Initial grades, though randomly assigned, did not affect the outcome as much. For C = 3, the range for SSE was 1486.36-1730.58, which is tighter range

than what we saw for K-Means, which was 1478.37-1915.95. The lowest SSE was 1486.36. Again, more variation in SSE is seen with increased number of clusters and initially assigned weights did affect the final centroids. C=5 had less variation in SSE than seen in C=3. SSE ranged from 669.86-783.03, which most ending with SSE around 783. Lowest SSE was 669.86, which was an outlier. The graphs shows that while the way clusters are grouped vary, SSE does not vary as much, which is very different from K-Means. C=7 showed more variation in SSE values. The SSE range was 515.91-623.47, which is still smaller range than what we saw in K-means, which was 489-644. Lowest SSE was 515.91. Most runs had SSE of around 516 or 524. As the graphs show, cluster formed varied between the runs but still managed to keep the SSE range small. K=10 had SSE range of 358.22-424.45, which is again smaller than what we have seen in K-Means, which was 333-468.The lowest SSE was at 358.22 but most SSE values were at 424.44.

**Conclusion**

SSE gets lower as the number of centroids, *c*, increases since it becomes easier to form tighter clusters as number of centroids increases. Since starting grades are randomly assigned, different clusters can form as centroid number increase. While the locations of final centroids vary, the SSE value does not vary as much for Fuzzy C-Means as it did for K-Means. Comparing the graphs of Fuzzy C-Means and K-Means shows similar centroid locations in many examples, but the borders of the clusters differ. For example, K=3 Run 4 graph and C=3 Run 3 graph has similar clusters and close enough SSE value, but you can see that Fuzzy C-Means determined that more points likely belong to the blue cluster (upper left cluster) near the border of the green cluster (middle cluster) than K-Means does for the light green cluster (upper left cluster) and the dark blue-green cluster (middle cluster). This difference is due to the fact that Fuzzy C-Means is more flexible in assigning points to clusters, and points can belong to more than one cluster. While the way I programmed the plotting colors assigned colored a point to reflect that of cluster with highest grade, those points actually can belong to more than one cluster, hence making the borders between the clusters weaker. This in turn seems to lower SSE and variability of SSE values.