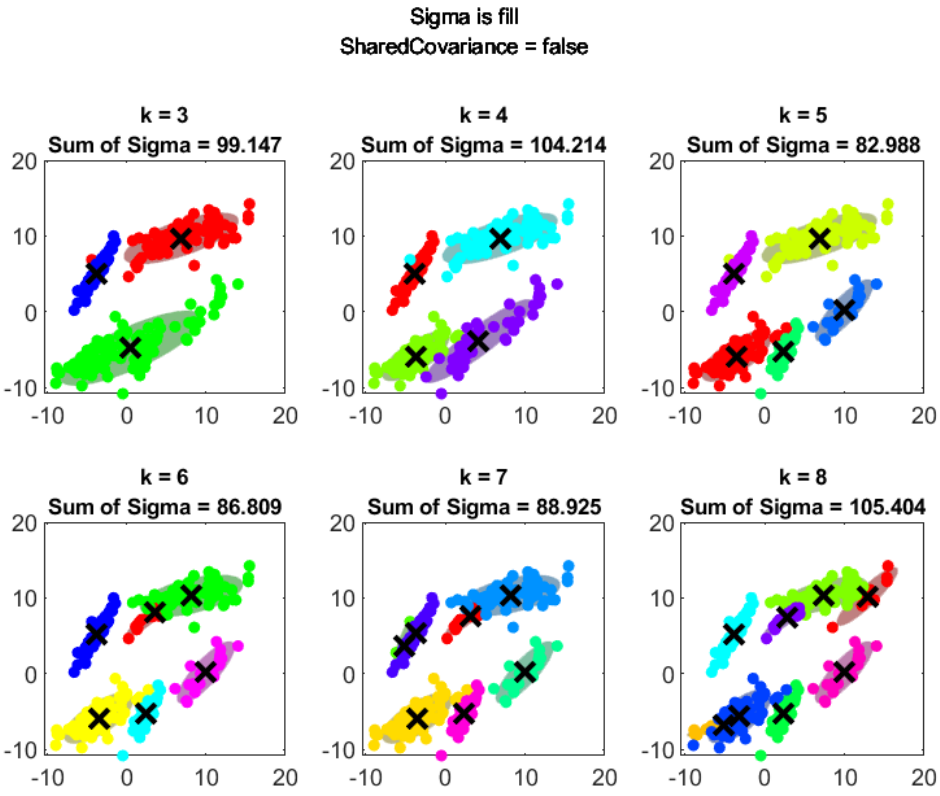
CS 565 / Assignment 2

Junehyeong Park

Oct 19, 2020

<Discussion on the question 2 of Assignment 2>



The K-Means method is the most easy and representative method for clustering. However, but there are unavoidable two disadvantages that the user will get random results due to specify the central point randomly so that it can occur the local optima, and how to select the number of clusters. To avoid two problems, I used GMM clustering method, and compared results of various number of clusters.

The GMM clustering method was applied with various number of clusters from 3 to 8. For the options, 'full' for the covariance matrix type and 'false' for the indicator for

identical or nonidentical covariance matrix were selected. The 'sigma', initial covariance matrix for all components, were summed and written on the titles of subplots. According to that, the number of clusters was selected as 5. Following table shows clustered points, written in the 'clustered.csv' result file.

1	4	5	3	6
2	10	7	11	8
17	20	9	22	12
26	48	14	31	13
29	64	15	38	19
34	70	16	39	32
40	76	18	43	37
42	107	21	54	41
47	143	23	59	44
49	149	24	62	52
51	170	25	63	53
55	176	27	66	61
58	192	28	80	65
60	228	30	84	71
67	239	33	86	85
69	245	35	88	87
73	249	36	93	91
77	262	45	109	94
78		46	113	96
79		50	121	97
89		56	122	98
92		57	131	101
99		68	141	105
102		72	146	111
103		74	147	112
104		75	150	114
117		81	152	115
118		82	154	127
119		83	156	128
120		90	164	129
123		95	166	138
125		100	167	145
130		106	182	148
132		108	185	162
134		110	186	165
135		116	190	173

140		124	194	174
142		126	199	175
151		133	200	178
157		136	206	180
158		137	209	181
159		139	220	187
168		144	221	188
169		153	223	193
171		155	224	201
184		160	230	203
198		161	231	207
202		163	246	210
204		172	251	211
208		177	267	212
213		179		217
214		183		219
218		189		237
222		191		240
226		195		241
232		196		244
233		197		248
235		205		250
236		215		254
242		216		256
243		225		257
259		227		258
260		229		
261		234		
263		238		
266		247		
		252		
		253		
		255		
		264		
		265		
		268		
		269		
		270		