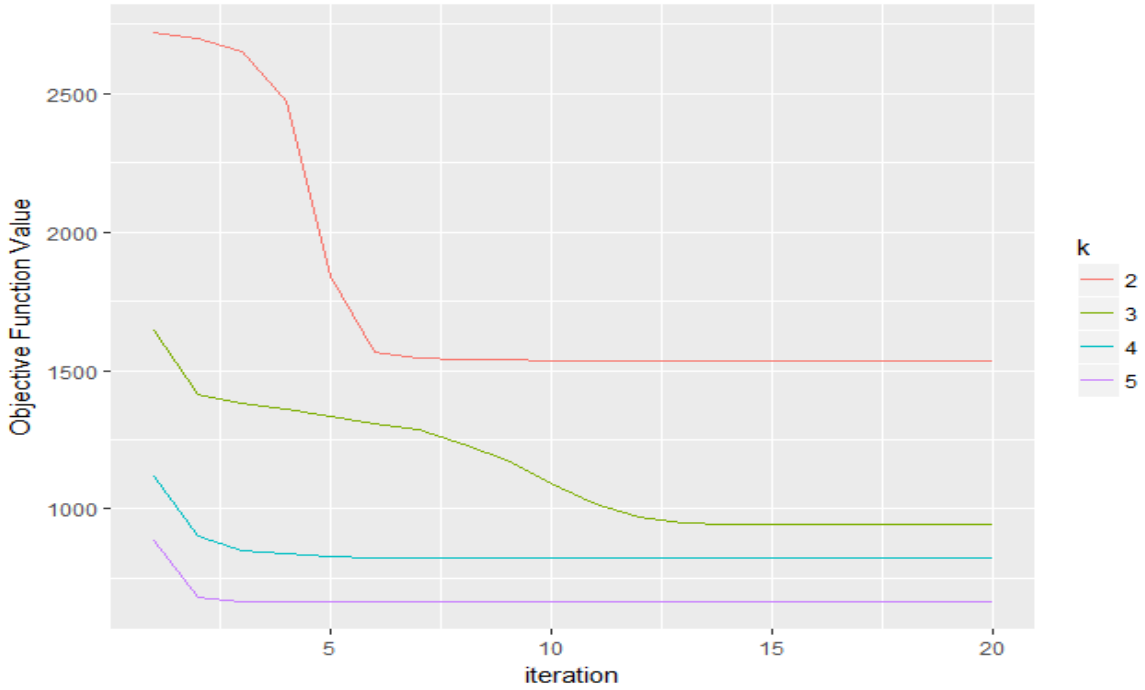
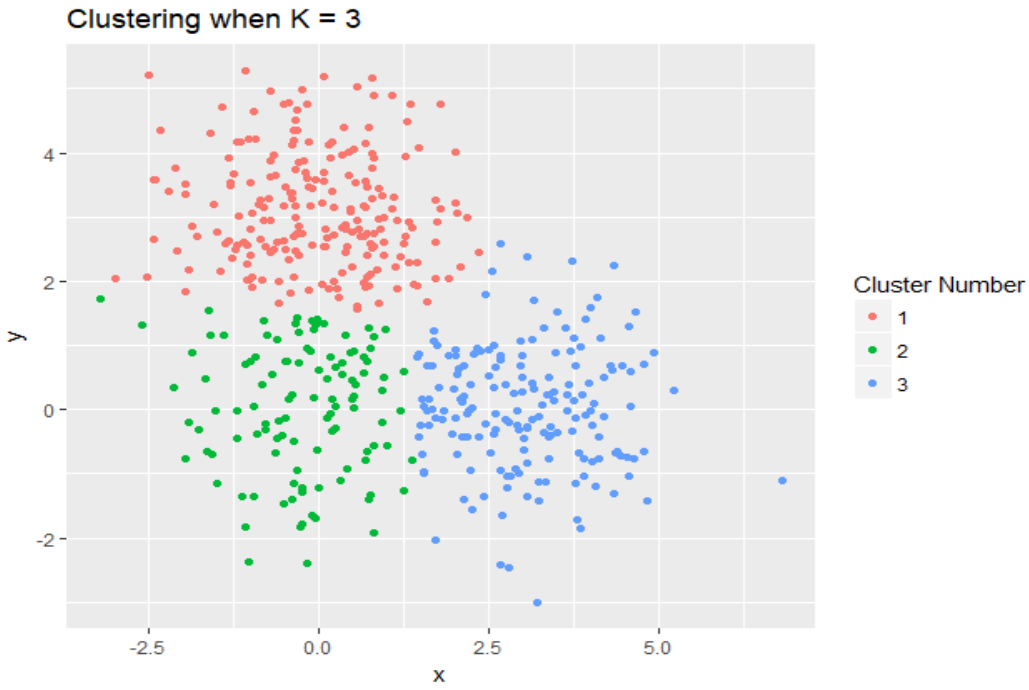
ELEN E4903: MACHINE LEARNING HOMEWORK 4

Problem 1:

a) After generating 500 observations from the mixture of the three Gaussians given in the question, the graph below depicts the values of the K-means objective function $\sum_{i=1}^{n}\sum_{k=1}^{K} 1\{c_i = k\}\|x_i - \mu_k\|^2$ per iteration for 20 iterations and for K = 2, 3, 4, and 5 and n = 500. As per the algorithm given in the lecture notes, c_i and μ_k were alternatingly updated. The results make sense since the objective function monotonically decreased (and eventually converged) as shown below.

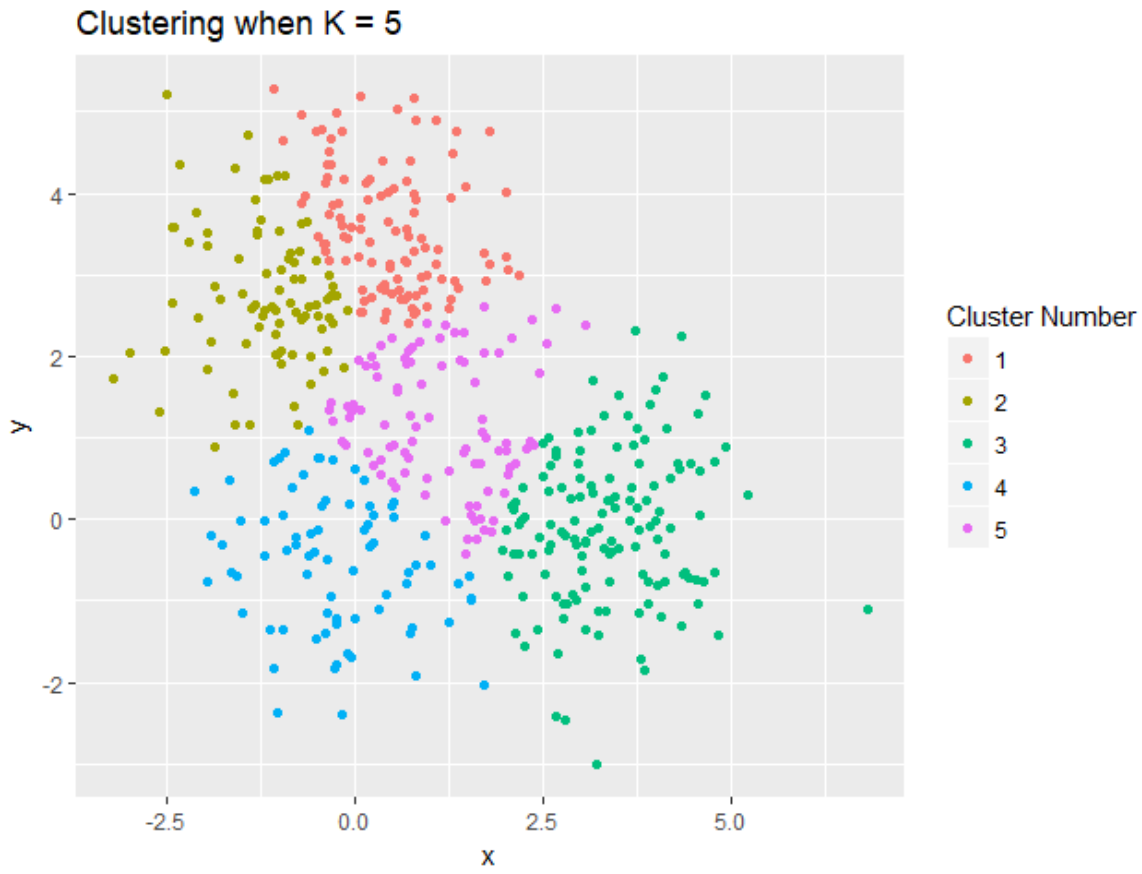


b) For K = 3, the following clusters are formed:



ELEN E4903: MACHINE LEARNING HOMEWORK 4

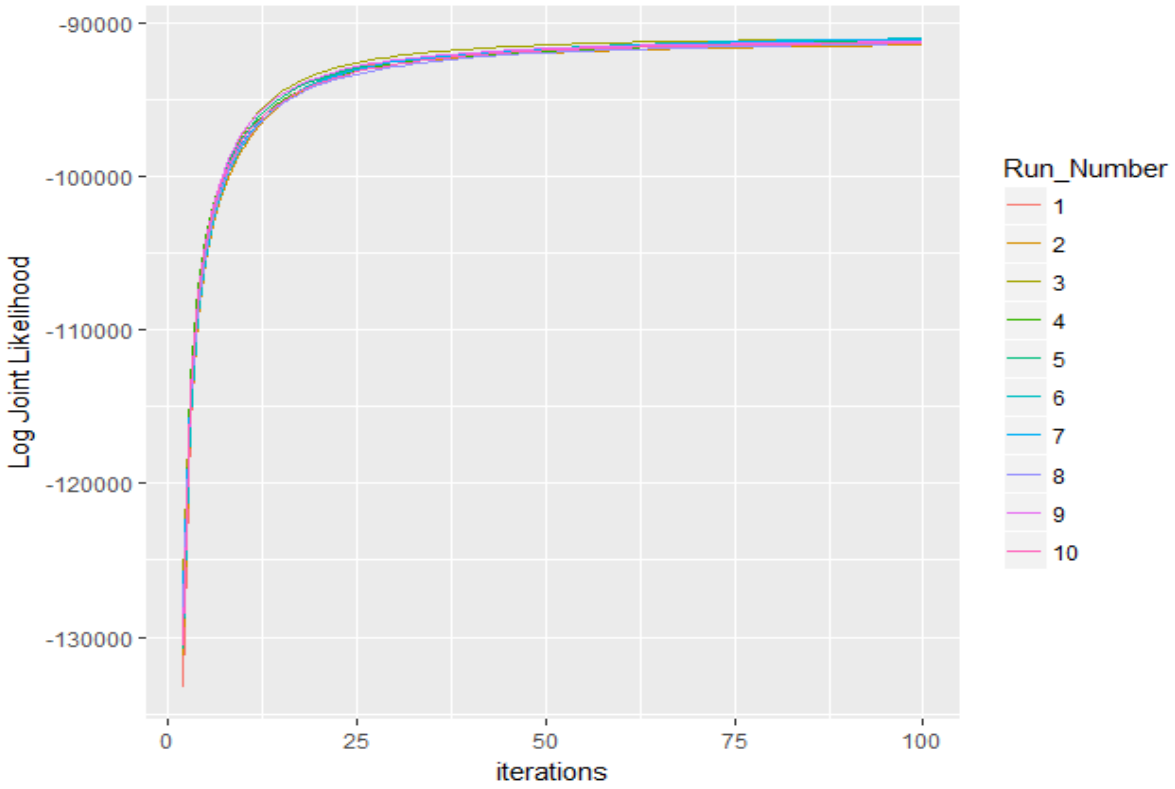
For K = 5, the following clusters are formed:



ELEN E4903: MACHINE LEARNING HOMEWORK 4

Problem 2:

a) The plot below shows the join log likelihood $-\sum_{(i,j)\in\Omega}\frac{1}{2\sigma^2}\left\|M_{ij}-u_i^T v_j\right\|^2-\sum_{i=1}^{N_1}\frac{\lambda}{2}\|u_i\|^2-\sum_{j=1}^{N_2}\frac{\lambda}{2}\left\|v_j\right\|^2$ for iterations 2 to 100 for each run of the code.



The table below shows the final value of the training objective function next to the RMSE on the testing set sorted according to decreasing value of the objective function.

Run No.	Final Training Obj Function Value	RMSE
6	-90961.59	1.286483
3	-90995.11	1.293108
7	-91008.62	1.281057
5	-91068.23	1.283544
1	-91119.48	1.280364
4	-91156.98	1.288703
9	-91172.00	1.290509
10	-91230.91	1.276821
8	-91282.21	1.280595
2	-91388.88	1.285527

ELEN E4903: MACHINE LEARNING HOMEWORK 4

b) The run with the highest objective value is 6. In the mapping from index to movie provided on the courseworks page, the query movie "Star Wars (1997)" is indexed as number 50, "My Fair Lady (1964)" is indexed as 485 while "Goodfellas (1990)" is indexed as 182. The following table shows the query movie, the ten nearest movies and their Euclidean distance.

"Star Wars (1997)" and its 10 closest movies:

movie_id	movie_name	Euclidean_distance
50	Star Wars (1977)	0.0000000
172	Empire Strikes Back, The (1980)	0.3535227
181	Return of the Jedi (1983)	0.6232150
174	Raiders of the Lost Ark (1981)	0.6813441
176	Aliens (1986)	0.9088293
194	Sting, The (1973)	0.9679254
12	Usual Suspects, The (1995)	0.9684726
210	Indiana Jones and the Last Crusade (1989)	0.9744278
195	Terminator, The (1984)	0.9905863
169	Wrong Trousers, The (1993)	1.0038735
173	Princess Bride, The (1987)	1.0050188

"My Fair Lady (1964)" and its 10 closest movies:

movie_id	movie_name	Euclidean_distance
485	My Fair Lady (1964)	0.0000000
419	Mary Poppins (1964)	0.6571813
99	Snow White and the Seven Dwarfs (1937)	0.7841397
133	Gone with the Wind (1939)	0.8199544
143	Sound of Music, The (1965)	0.8701294
1147	My Family (1995)	0.8912193
606	All About Eve (1950)	0.9065276
417	Parent Trap, The (1961)	0.9336028
604	It Happened One Night (1934)	0.9422090
418	Cinderella (1950)	0.9538514
633	Christmas Carol, A (1938)	0.9578888

ELEN E4903: MACHINE LEARNING HOMEWORK 4

"GoodFellas (1990)" and its 10 closest movies:

movie_id	movie_name	Euclidean_distance
182	GoodFellas (1990)	0.0000000
693	Casino (1995)	0.6531517
177	Good, The Bad and The Ugly, The (1966)	0.7523679
187	Godfather: Part II, The (1974)	0.7793538
504	Bonnie and Clyde (1967)	0.8487971
523	Cool Hand Luke (1967)	0.8792199
33	Desperado (1995)	0.8800824
205	Patton (1970)	0.9348605
76	Carlito's Way (1993)	0.9602970
180	Apocalypse Now (1979)	0.9672324
188	Full Metal Jacket (1987)	0.9930162