Problem 1 (K-means):

a) For K = 2; 3; 4; 5, plot the value of the K-means objective function per iteration for 20 iterations (the algorithm may converge before that).
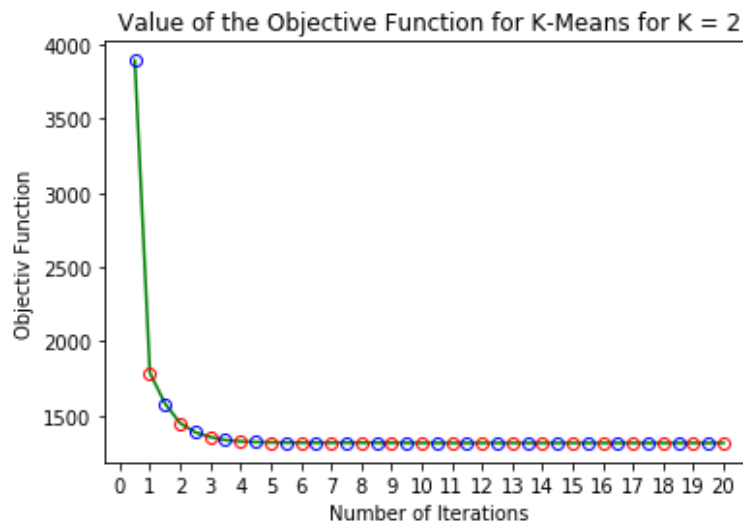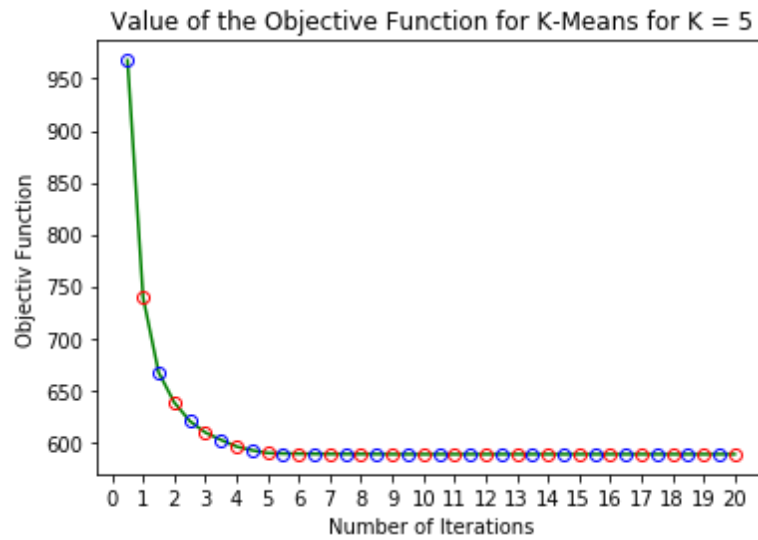

Value of the Objective Function for K-Means for K = 2

Value of the Objective Function for K-Means for K = 3

Value of the Objective Function for K-Means for K = 4

Value of the Objective Function for K-Means for K = 5
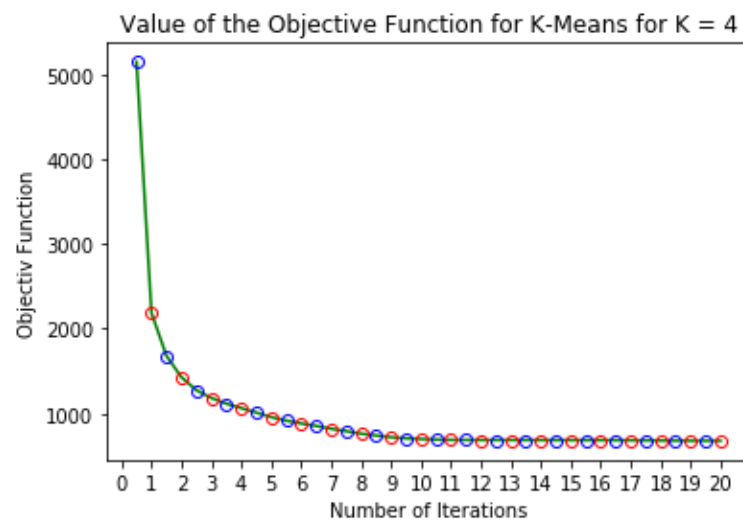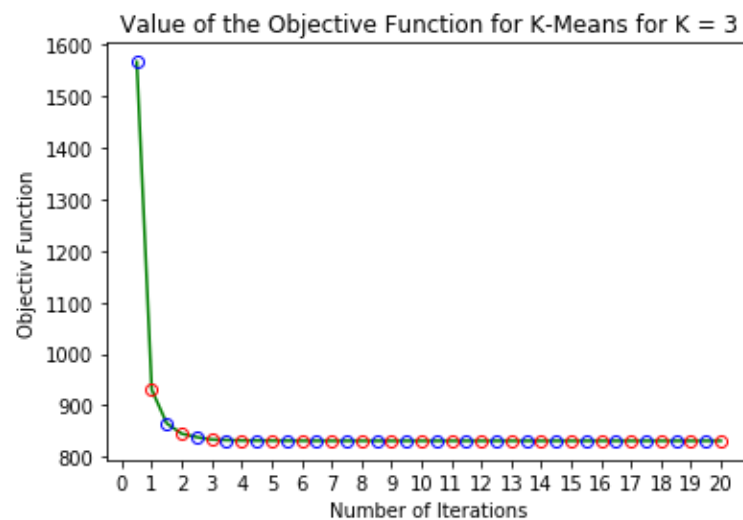
b) For K = 3; 5, plot the 500 data points and indicate the cluster of each for the final iteration by marking it with a colour or a symbol.

- When K = 3:
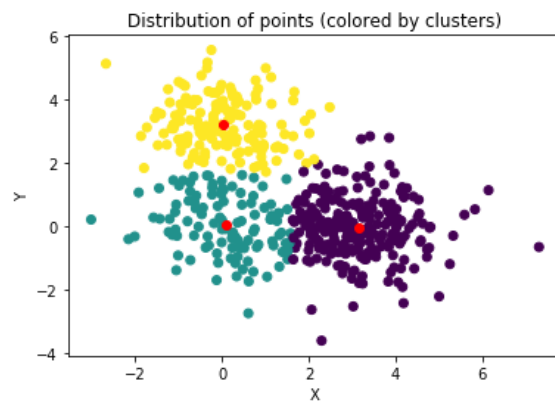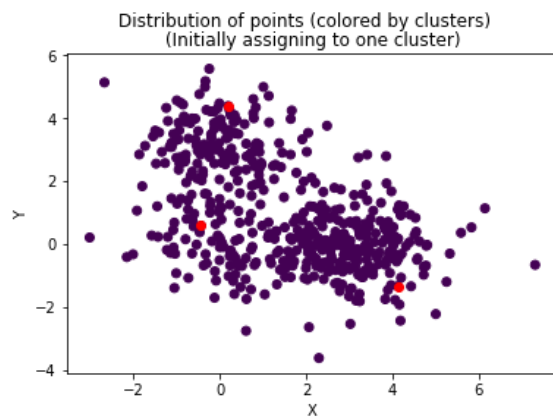
- When K = 5:
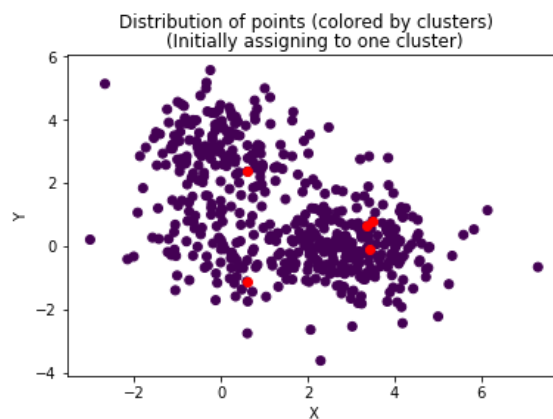
Problem 2 (Bayes classifier revisited):

a) Implement the EM algorithm for the GMM described in class. Using the training data provided, for each class separately, plot the log marginal objective function for a 3-Gaussian mixture model over 10 different runs and for iterations 5 to 30. There should be two plots, each with 10 curves.
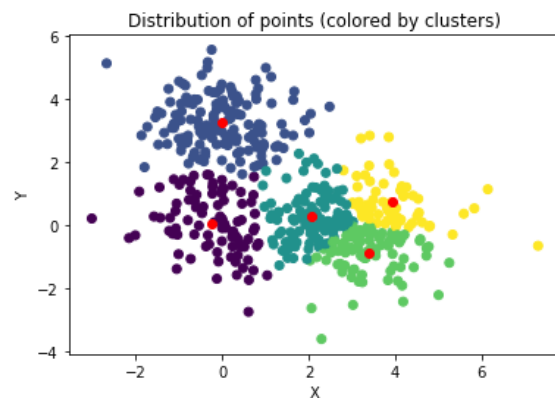
- Following figure shows the plot of log marginal objective function for a 3- Gaussian Mixture Model over 10 different runs for iterations 5-30.

For Class: 0

Log marginal objective function for a 3-Gaussian mixture model over 10 different runs and for iterations 5 to 30

For Class: 1

Log marginal objective function for a 3-Gaussian mixture model over 10 different runs and for iterations 5 to 30
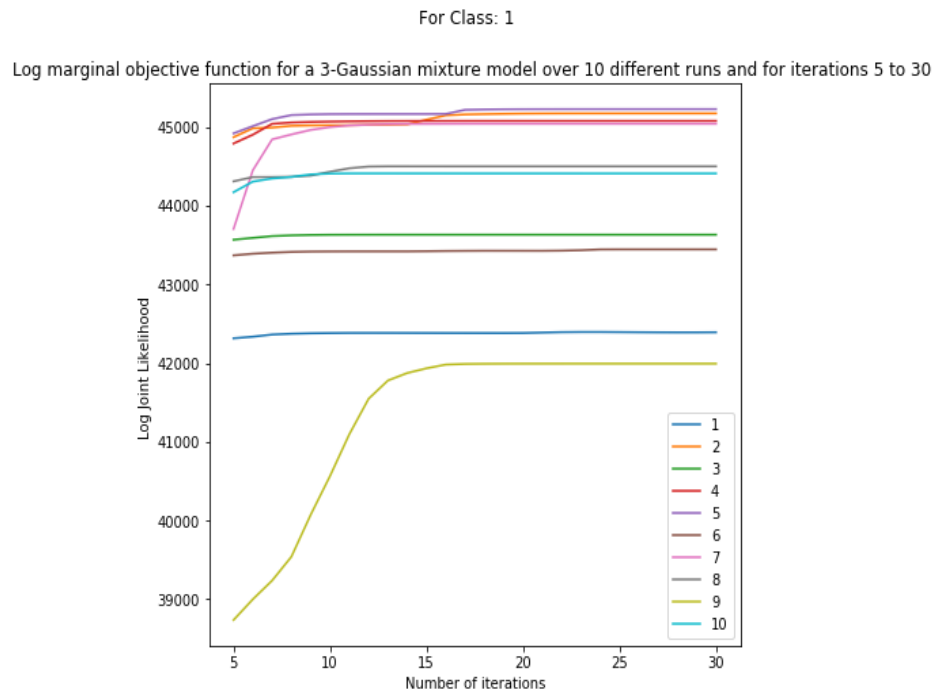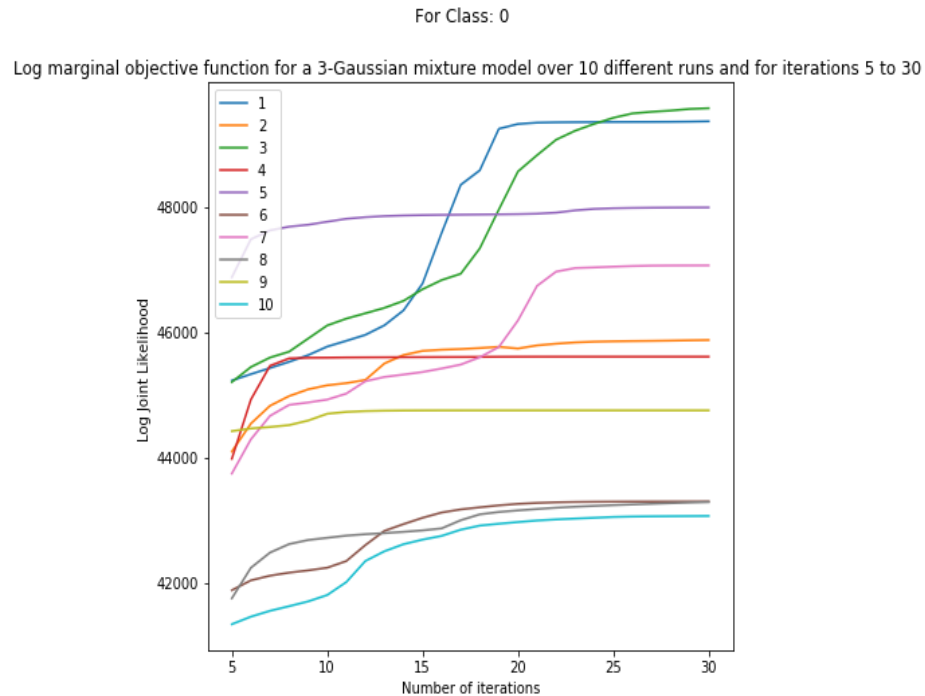
b) Using the best run for each class after 30 iterations, predict the testing data using a Bayes classifier and show the result in a 2 x 2 confusion matrix, along with the accuracy percentage. Repeat this process for a 1-, 2-, 3- and 4-Gaussian mixture model. Show all results nearby each other, and don't repeat Part (a) for these other cases. Note that a 1-Gaussian GMM doesn't require an algorithm, although your implementation will likely still work in this case.

```
The results for 1- Gaussian Mixture Model
      0    1
0   182   96
1     7  175
Accuracy: 0.78
```

```
The results for 2- Gaussian Mixture Model
      0    1
0   198   80
1     9  173
Accuracy: 0.81
```

```
The results for 3- Gaussian Mixture Model
      0    1
0   220   58
1     9  173
Accuracy: 0.85
```

```
The results for 4- Gaussian Mixture Model
      0    1
0   206   72
1    15  167
Accuracy: 0.81
```

Problem 3 (Matrix factorization):

a) Run your code 10 times. For each run, initialize your u_i and v_j vectors as N(0; I) random vectors. On a single plot, show the the log joint likelihood for iterations 2 to 100 for each run. In a table, show in each row the final value of the training objective function next to the RMSE on the testing set. Sort these rows according to decreasing value of the objective function.

- Following figure shows the plot of the loglikelihood for iterations 2 to 100 for each of the 10 runs
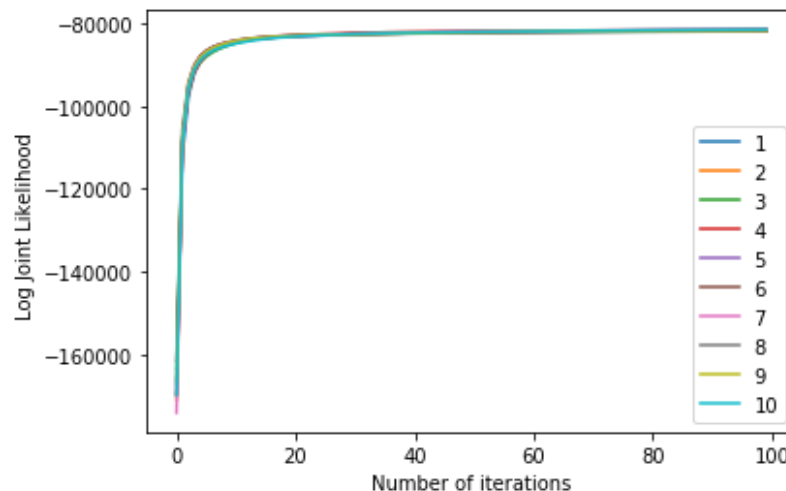
- The following table shows the RMSE on the testing set for each of the run. (The values are sorted by decreasing value of objective function)

Iteration	Likelihood	RMSE
9	-81549.665049	1.111169
0	-81567.959526	1.104616
3	-81589.365281	1.105367
6	-81616.382220	1.115418
4	-81740.272567	1.104373

Iteration	Likelihood	RMSE
8	-81756.976140	1.100690
5	-81765.518308	1.123615
1	-81865.093456	1.089419
2	-81917.010864	1.096269
7	-81958.337459	1.092063

b) For the run with the highest objective value, pick the movies "Star Wars" "My Fair Lady" and "Goodfellas" and for each movie find the 10 closest movies according to Euclidean distance using their respective locations v_j. List the query movie, the ten nearest movies and their distances. A mapping from index to movie is provided with the data.

```
-----------------------------------------------------------------
Similar movies for Star Wars (1977)
                                            Movies   Distances
0                                Delicatessen (1991)   0.097009
1                          Princess Bride, The (1987)   0.388544
2                           Apocalypse Now (1979)      0.416990
3                         Harold and Maude (1971)      0.698186
4                    Empire Strikes Back, The (1980)   0.741952
5                                      Psycho (1960)   0.811132
6                          Jean de Florette (1986)     0.814723
7   Star Trek VI: The Undiscovered Country (1991)     0.900430
8                           Right Stuff, The (1983)    0.922538
9                             Down by Law (1986)       0.949635
-----------------------------------------------------------------
```

```
-----------------------------------------------------------------
Similar movies for My Fair Lady (1964)
                         Movies   Distances
0     Gone with the Wind (1939)   0.644885
1          Mary Poppins (1964)    1.119822
2            Funny Face (1957)    1.150138
3                  Babe (1995)    1.391833
4       Parent Trap, The (1961)   1.558116
5          Philadelphia (1993)    1.568831
6              Sleepers (1996)    1.670807
7       Wizard of Oz, The (1939)  1.722190
8          Fly Away Home (1996)   1.732144
9   Alice in Wonderland (1951)    1.759021
-----------------------------------------------------------------
```

```
-----------------------------------------------------------------
Similar movies for GoodFellas (1990)
                                       Movies   Distances
0            Godfather: Part II, The (1974)     0.251683
1                     Candidate, The (1972)     0.393383
2   Good, The Bad and The Ugly, The (1966)     0.509839
3             American President, The (1995)    0.518856
4               2001: A Space Odyssey (1968)    0.642817
5                 Blues Brothers, The (1980)    0.704650
6                       Godfather, The (1972)   0.797259
7                             Platoon (1986)    0.798328
8                     Apocalypse Now (1979)     0.900091
9                       Down by Law (1986)      0.906754
```