Francisco Javier Arceo
fja2114@columbia.edu
EECS6892: Bayesian Machine Learning
Homework 4

Problem 1.

a. See code attached in EECS6892_HMWK4_FJA.py and refer to:

- **EM_GMM()**
- **llkplot()**
- **em_scatplot()**

b. Refer to plots attached

The log-likelihood strictly increases for each k across the iterations. More importantly, the log-likelihood continues to increase as K grows larger. This would likely lead to overfitting clusters on the data, I suspect that the log-likelihood would continue to increase as a function of K, so we would be led to believe that the optimal model would be the one with the largest K, which would not be correct.

c. Refer to plots attached

As a function of K, these plots select smaller partitions of the data in 2-d. As K gets larger, the clusters become less intuitive; for example, the plot with 10 clusters has a strange behavior for the orange cluster. This is because with EM we can always find the best cluster given the mean, but that doesn't necessarily imply that the clusters themselves are generalizable to other data.

Problem 2.

a. See code attached in EECS6892_HMWK4_FJA.py and refer to:

- **VI_GMM()**
- **vof_plot()**
- **vi_scatplot()**

b. Refer to plots attached

Unlike EM, the Variational Objective function is not monotonically increasing, which means (either there's a bug in my code) or it's not as stable as EM. More importantly, just like the log-likelihood in EM, the VI seems to prefer larger K, yet, strangely, the variational objective function for K = 10 is larger than for K = 25

c. Refer to plots attached

Several things are worth noting about the VI implementation and the resulting clusters. (1) It is very sensitive to initialization. I iterated through 50 random seeds to verify this and I found that the number of clusters when k=4 varied quite dramatically; (2) it can often lead to clusters < K, for example, the model with K = 10 only yields 4 clusters and the model with K = 25 only yields 3 clusters, which is very unexpected. In some cases, I was able to recover 8 clusters or more, but this was a rarity.

Problem 3.

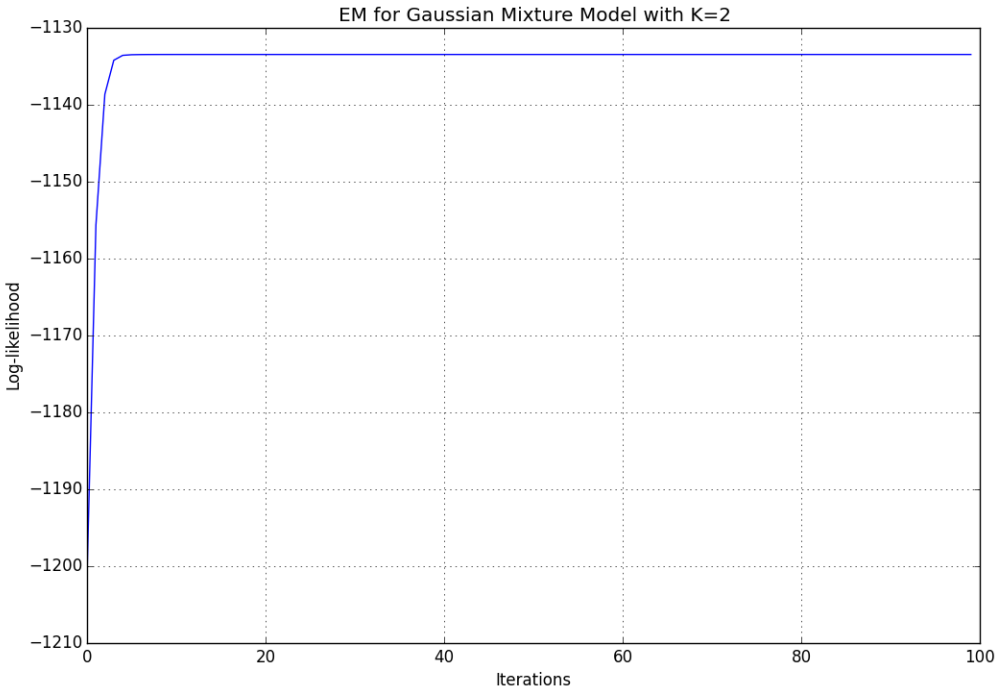
a. See code attached in EECS6892_HMWK4_FJA.py refer to following functions:

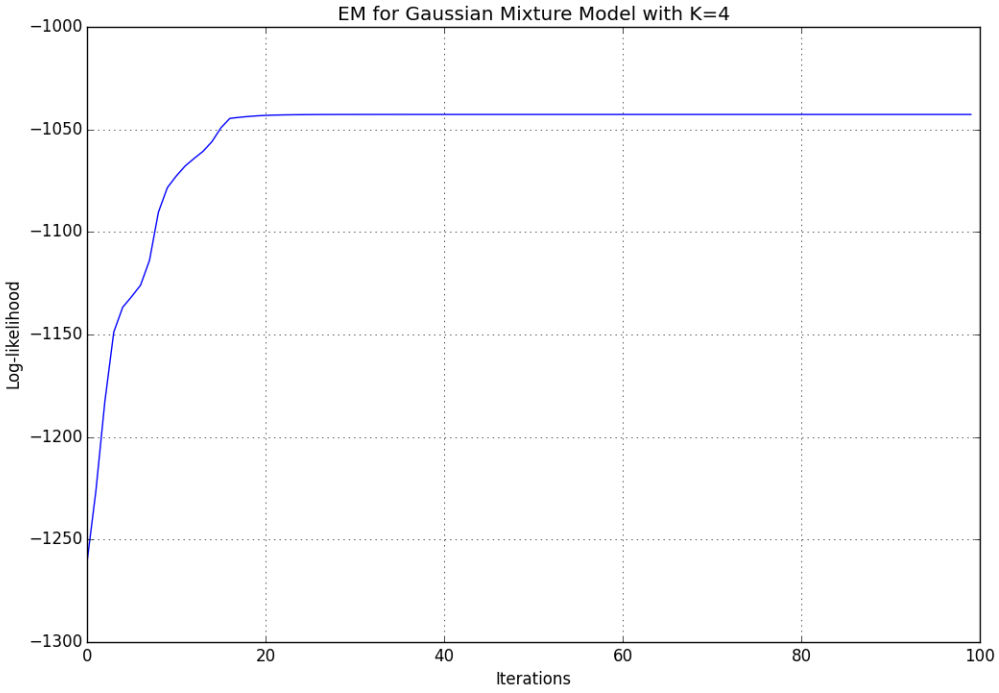
- **GMM_Gibbs_Sampler()**
- **update_posterior()**
- **calculate_px()**
- **plot_ClusterCount ()**
- **plot_TotalClusters ()**

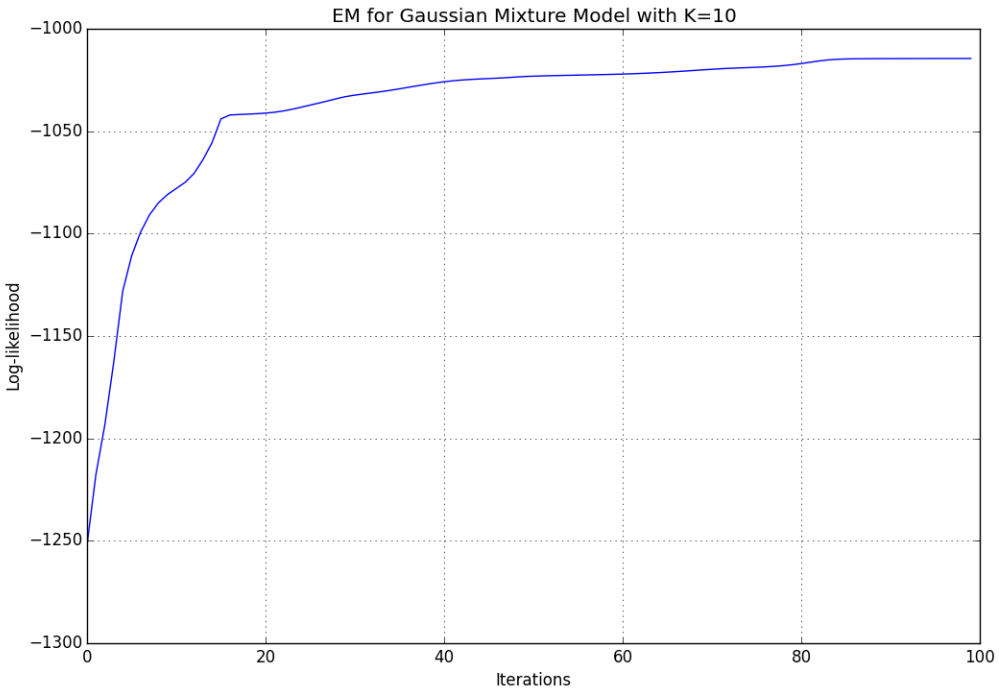
b. Refer to plots attached

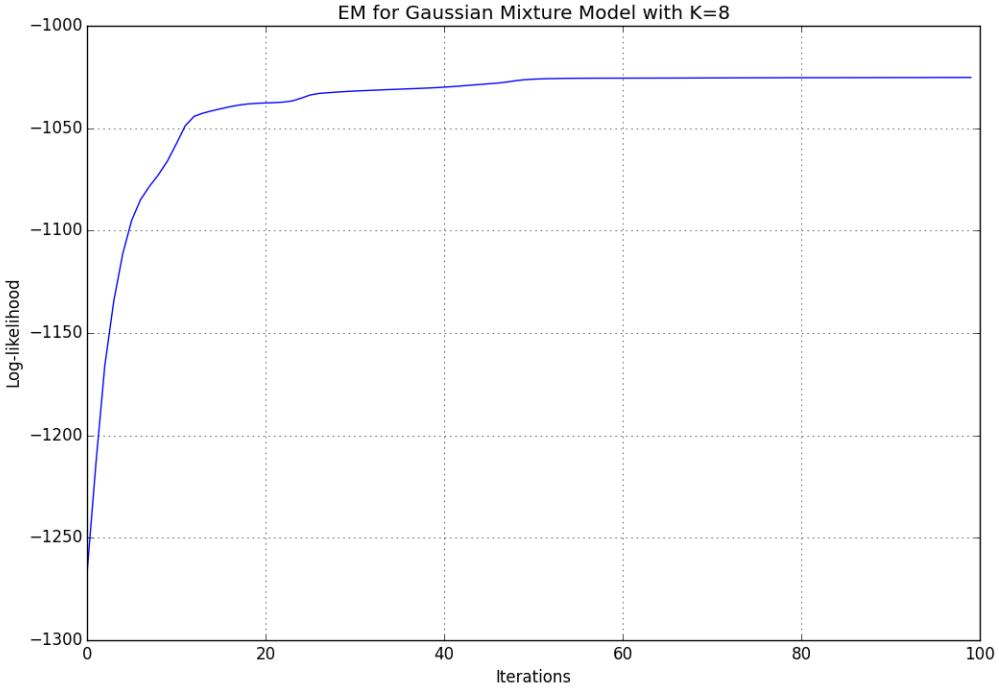
c. Refer to plots attached

Plots for Problem 1.b.

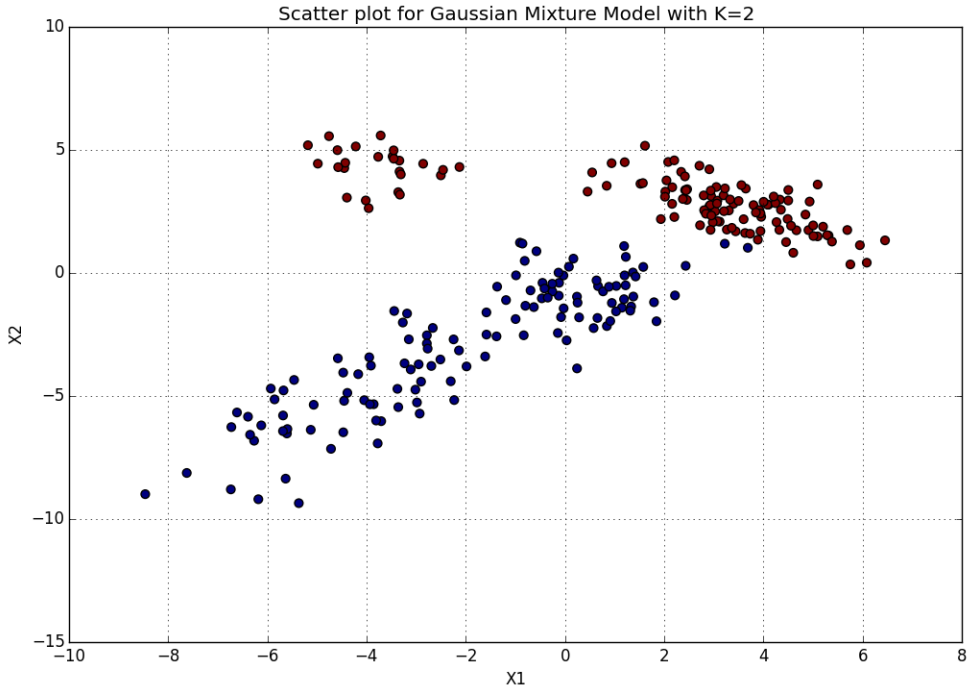


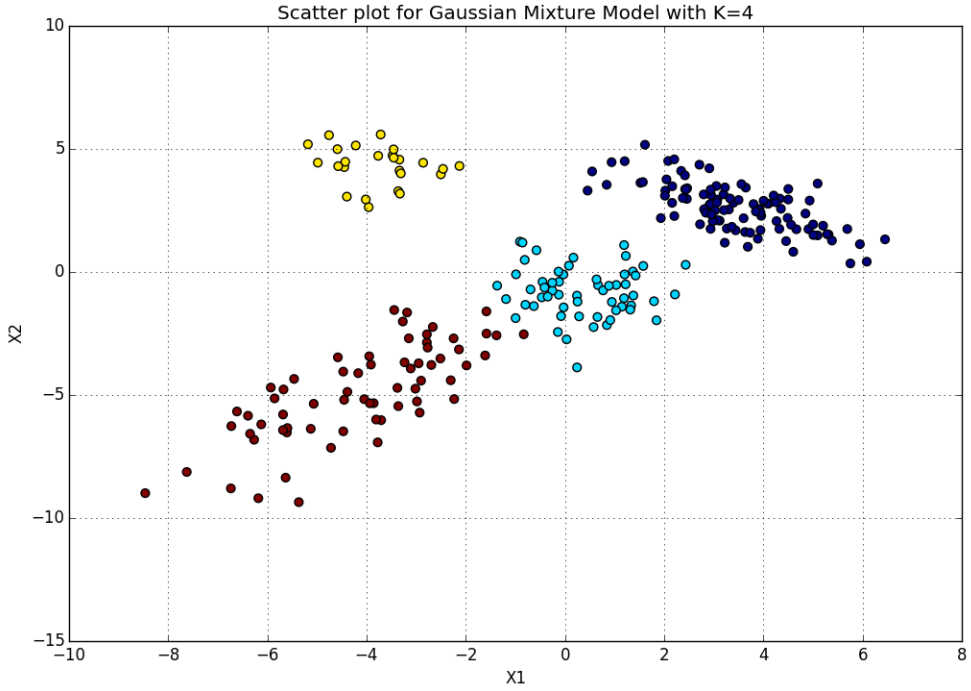


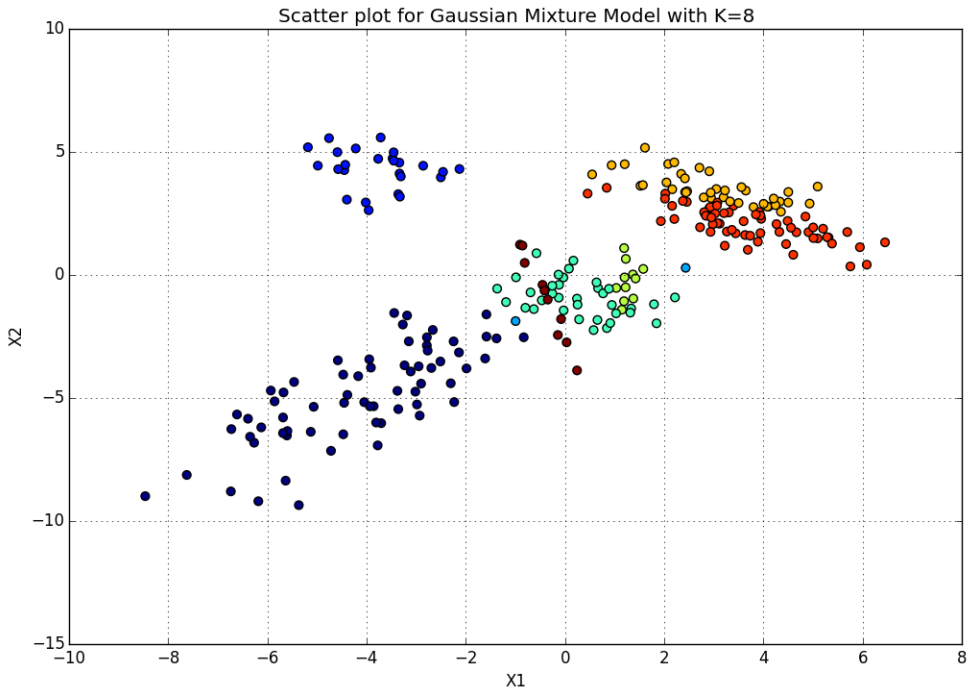

EM for Gaussian Mixture Model with K=10
−1000
−1050
−1100
−1150
−1200
−1250
−1300
Log-likelihood
0
20
40
60
80
100
Iterations

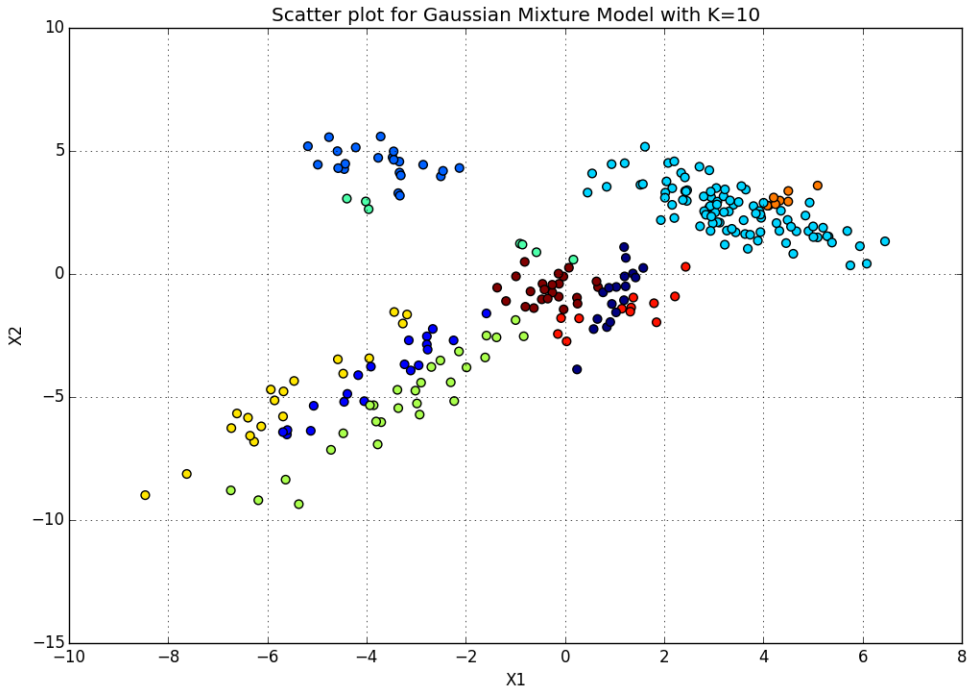

EM for Gaussian Mixture Model with K=8
−1000
−1050
−1100
−1150
−1200
−1250
−1300
Log-likelihood
0
20
40
60
80
100
Iterations

Plots for Problem 1.c.

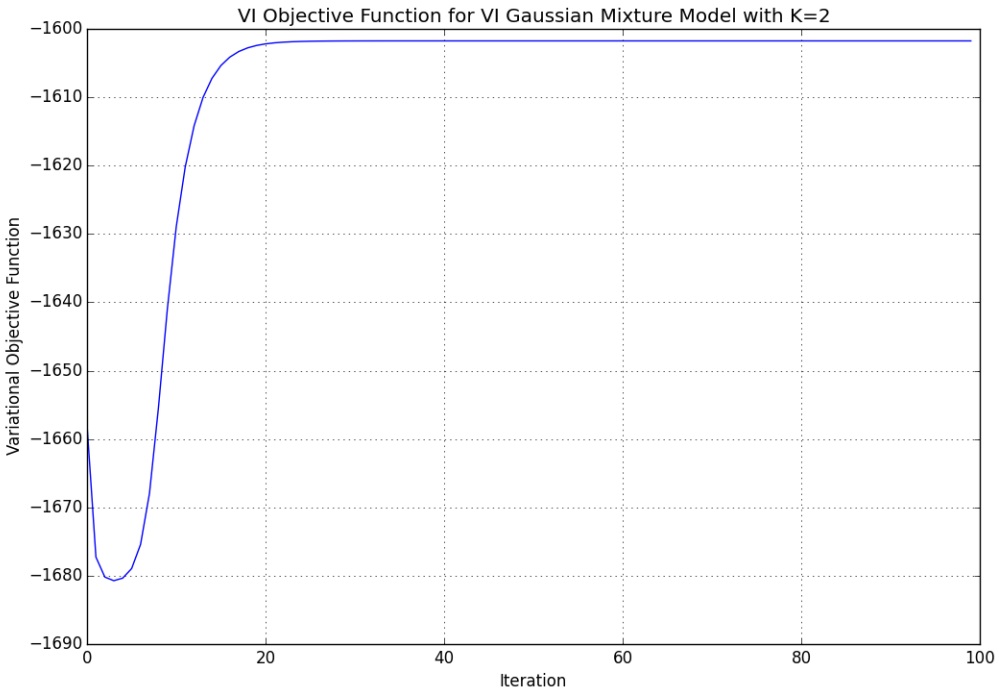


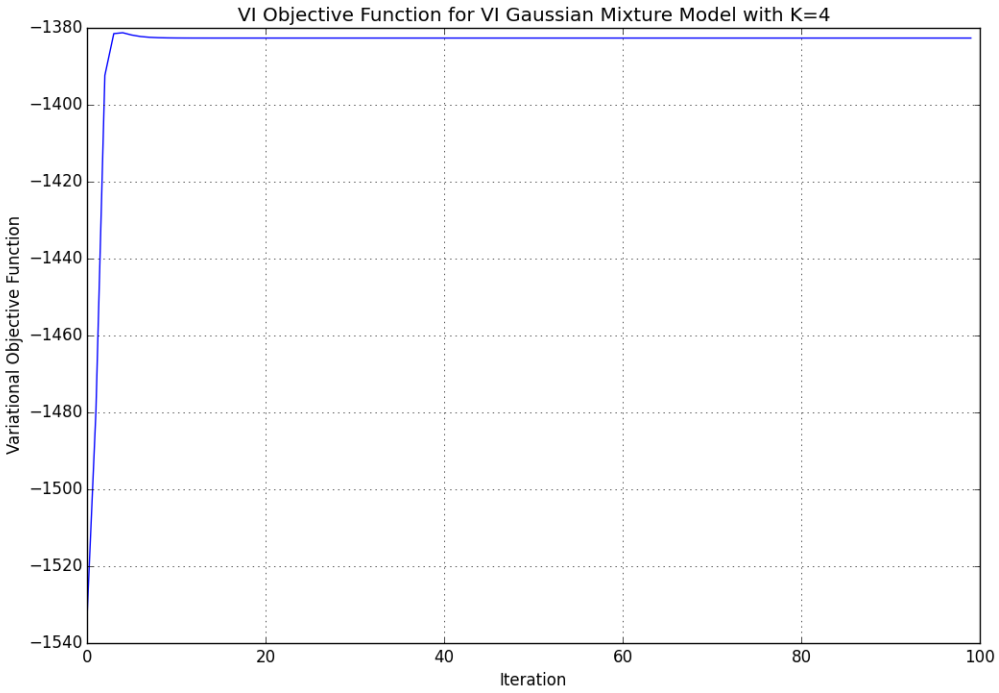


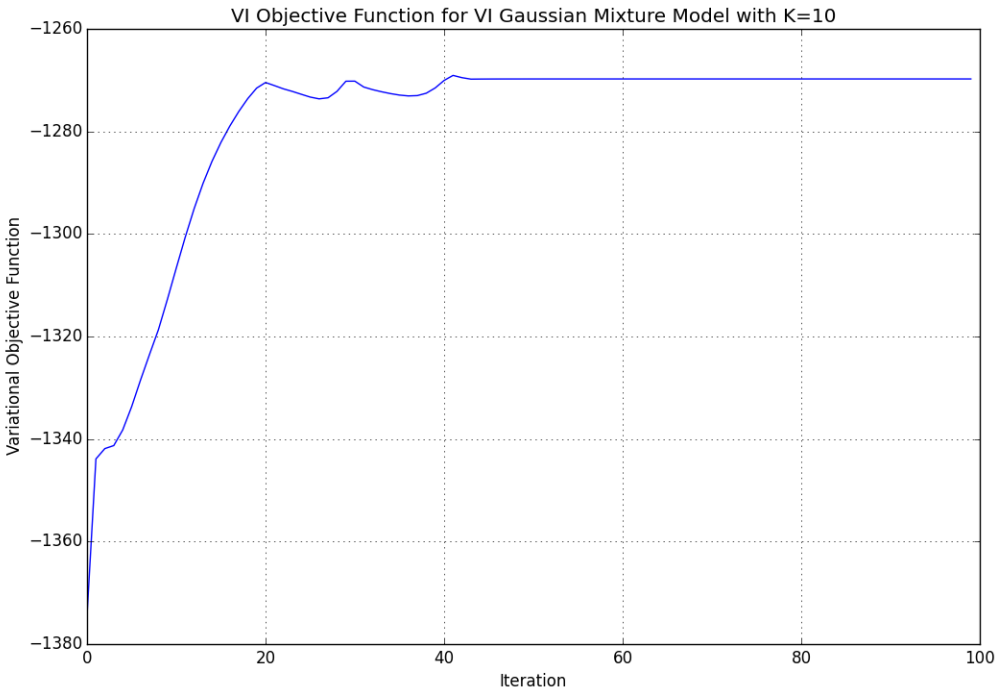

Scatter plot for Gaussian Mixture Model with K=8
X1
X2
10
5
0
−5
−10
−15
−10
−8
−6
−4
−2
0
2
4
6
8

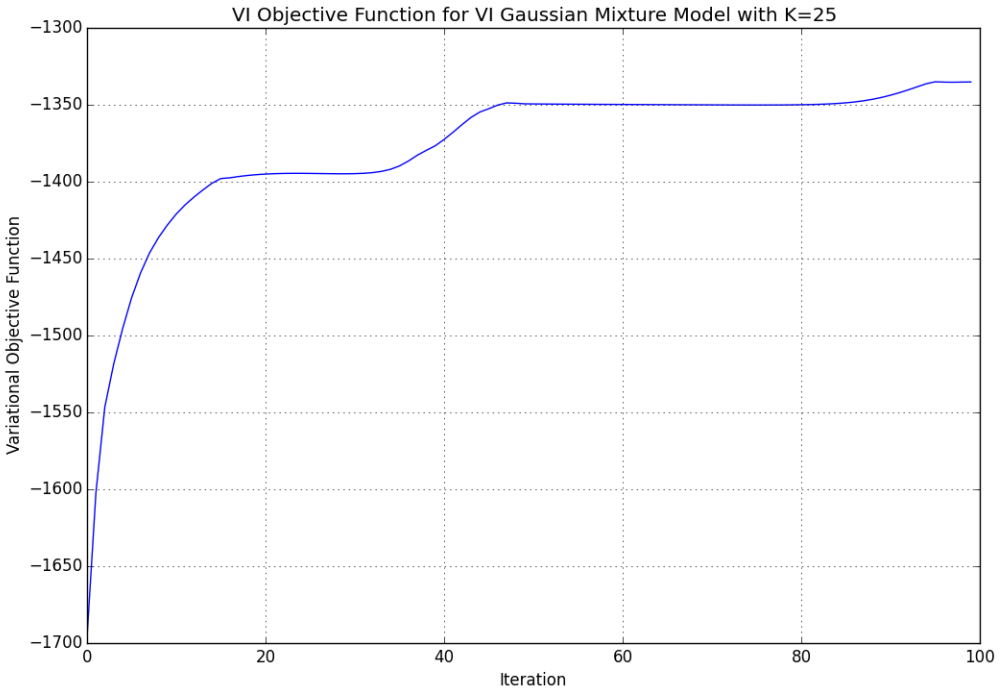

Scatter plot for Gaussian Mixture Model with K=10
X1
X2
10
5
0
−5
−10
−15
−10
−8
−6
−4
−2
0
2
4
6
8

Plots for Problem 2.b.

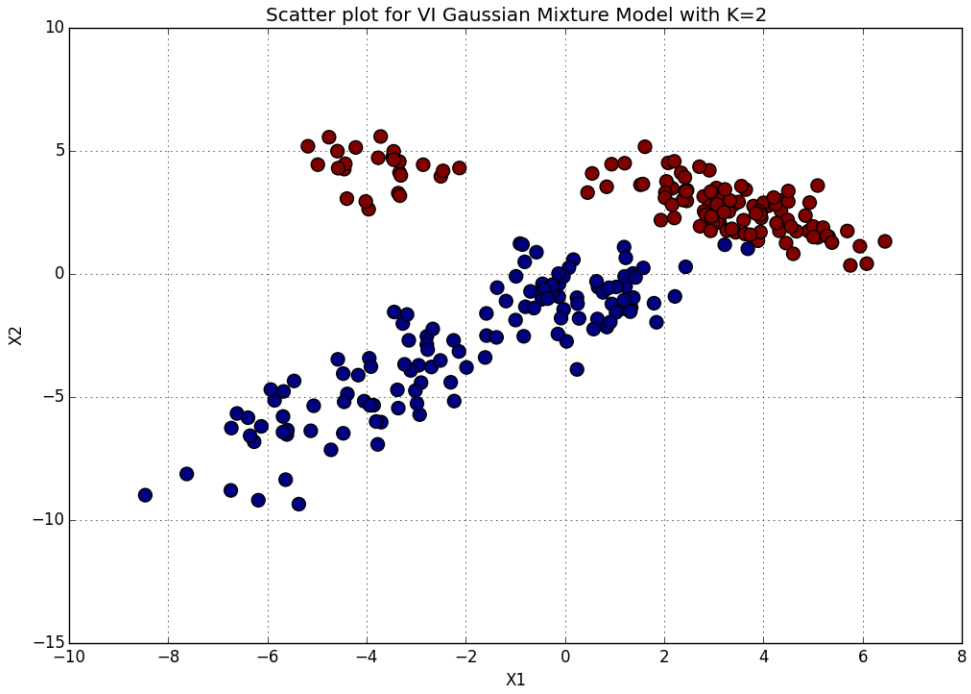


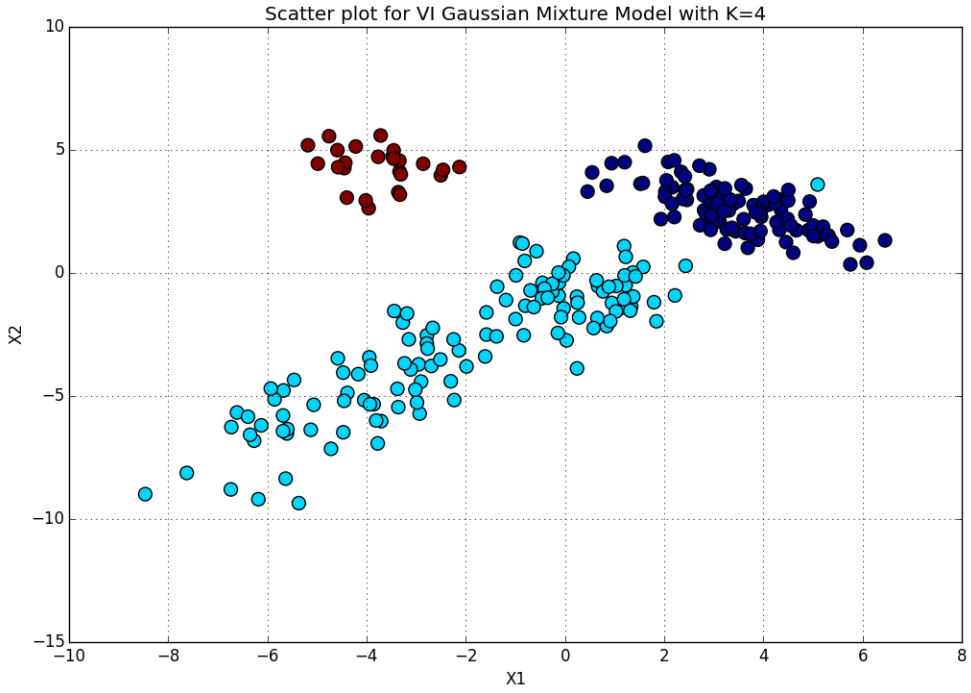


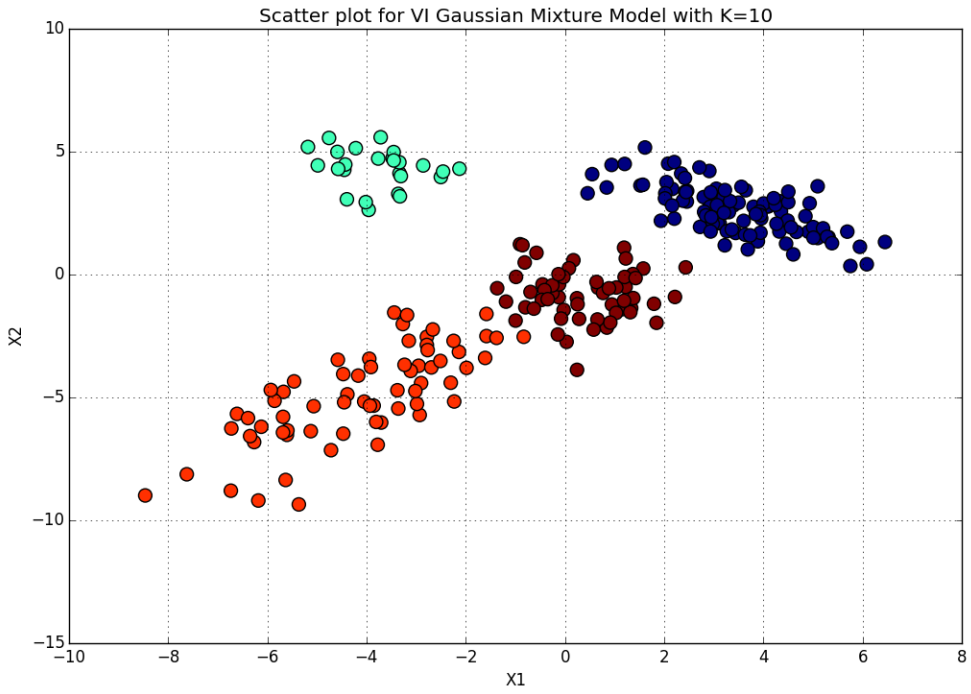

VI Objective Function for VI Gaussian Mixture Model with K=10
Variational Objective Function
Iteration
−1260
−1280
−1300
−1320
−1340
−1360
−1380
0
20
40
60
80
100

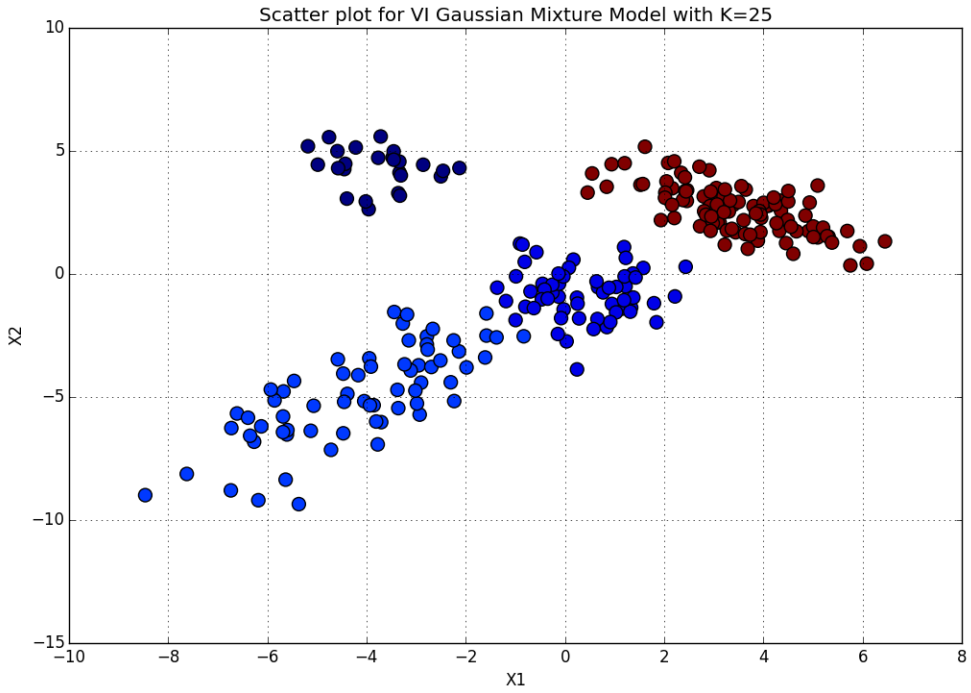

VI Objective Function for VI Gaussian Mixture Model with K=25
Variational Objective Function
Iteration
−1300
−1350
−1400
−1450
−1500
−1550
−1600
−1650
−1700
0
20
40
60
80
100

Plots for Problem 2.c.

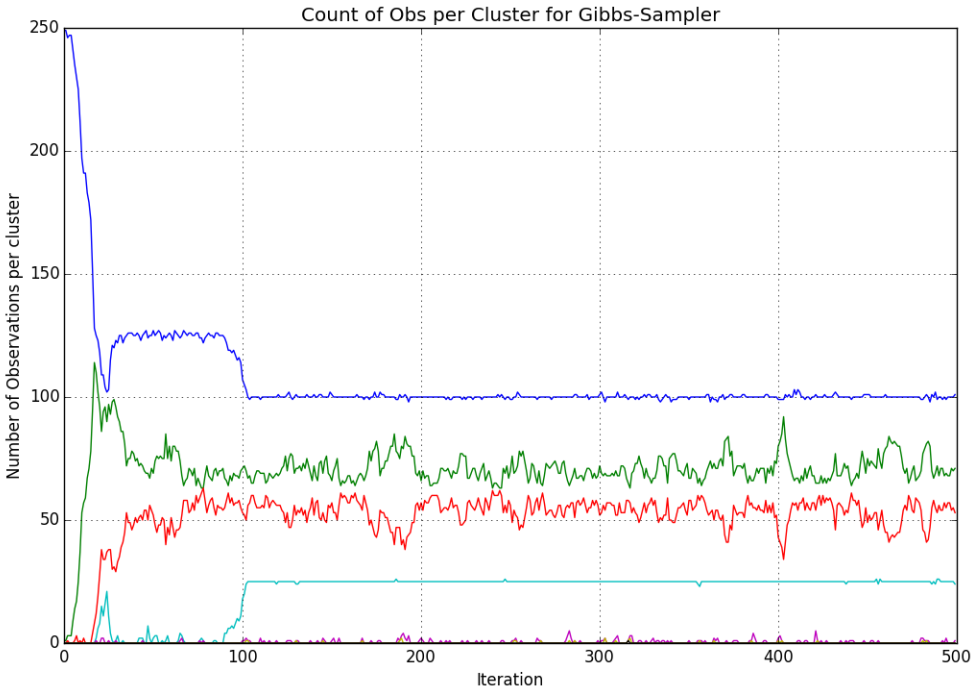





Scatter plot for VI Gaussian Mixture Model with K=10
X1
X2


Scatter plot for VI Gaussian Mixture Model with K=25
X1
X2

Plots for Problem 3.b.



Plots for Problem 3.c.