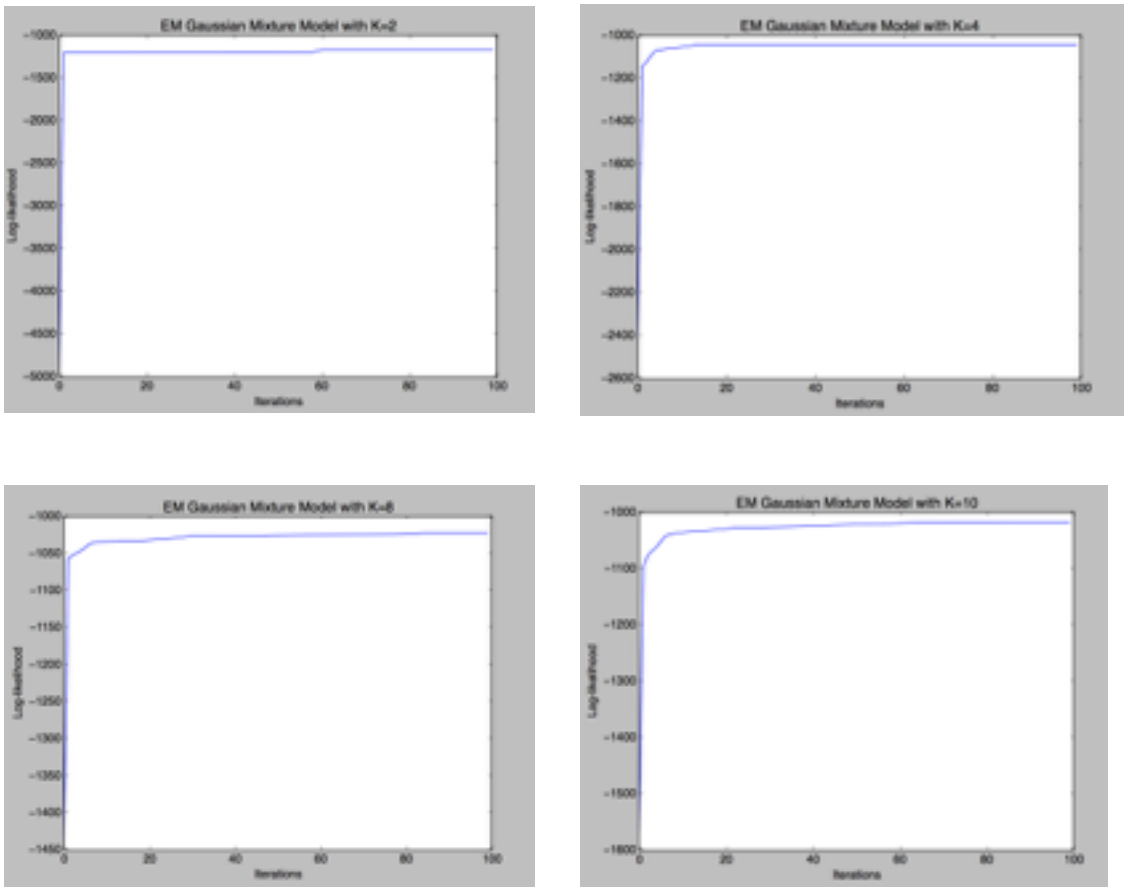
Homework 4

Problem 1:

a) **Implement the EM-GMM algorithm and run it for 100 iterations on the data provided for K = 2,4,8,10.**

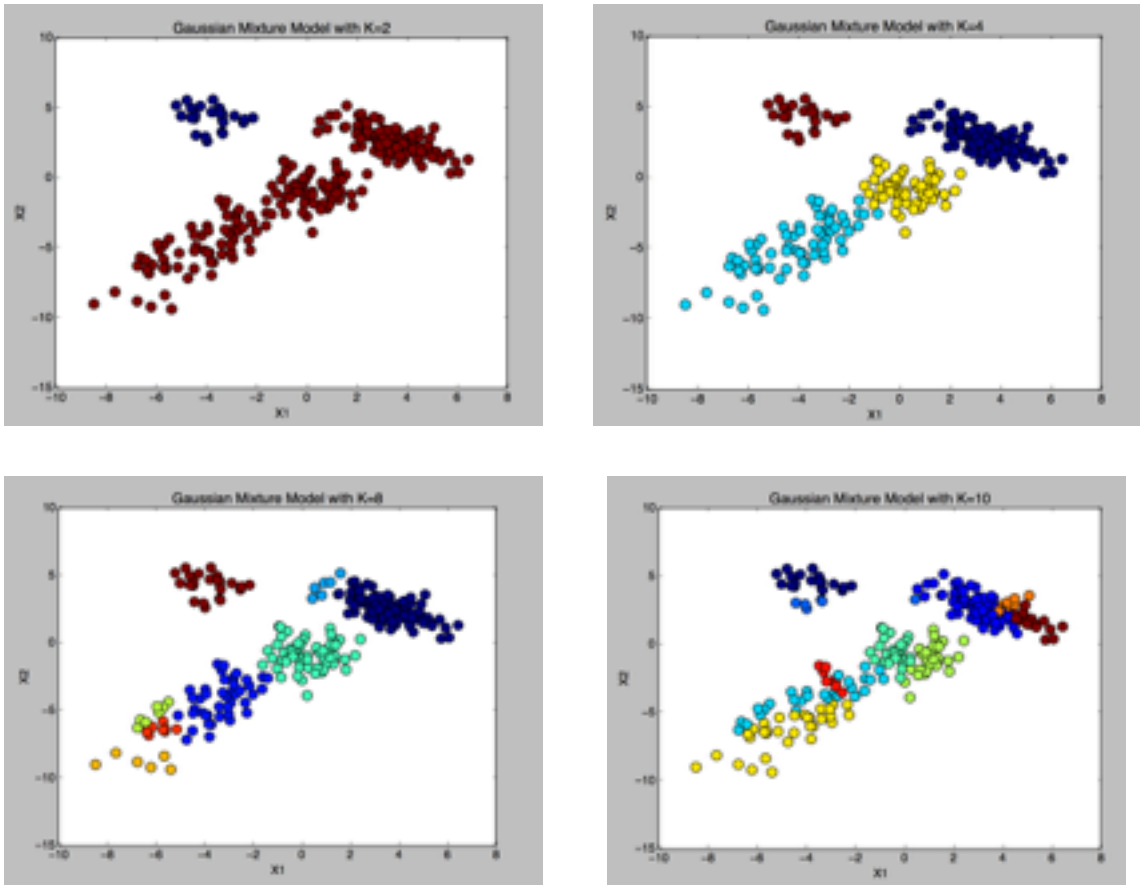
See code attached.

b) **For each K, plot the log likelihood over the 100 iterations. What pattern do you observe and why might this not be the best way to do model selection?**



For all k's, the log-likelihood increases. Even when we can see that the clusters are way overfitting the data the log-likelihood continues to increase.

c) For the final iteration of each model, plot the data and indicate the most probable cluster of each observation according to $q(c_i)$ by a cluster-specific symbol. What do you notice about these plots as a function of K?



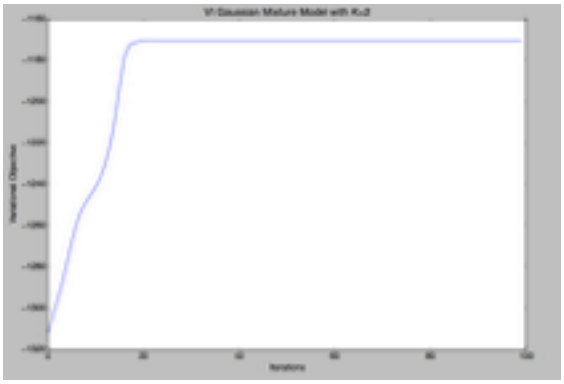
We can see that as K becomes larger and larger, the behavior becomes less intuitive. You can sometimes see weird things where there is one dot on the other side of a cluster and I believe that this can be due to the large variance on certain gaussians.

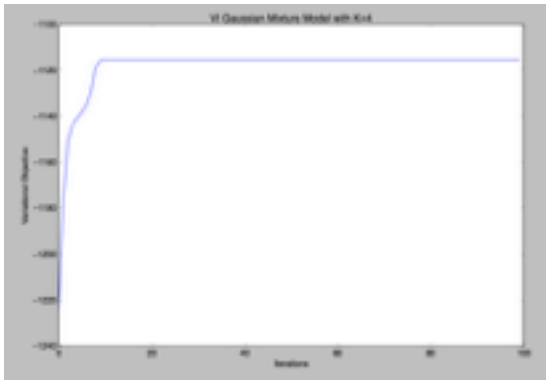
Problem 2:

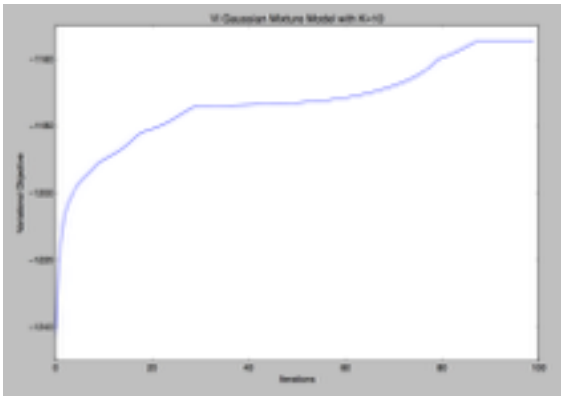
a) Implement the variational inference algorithm discussed in class and in the notes for K = 2, 4, 10, 25 and 100 iterations each.

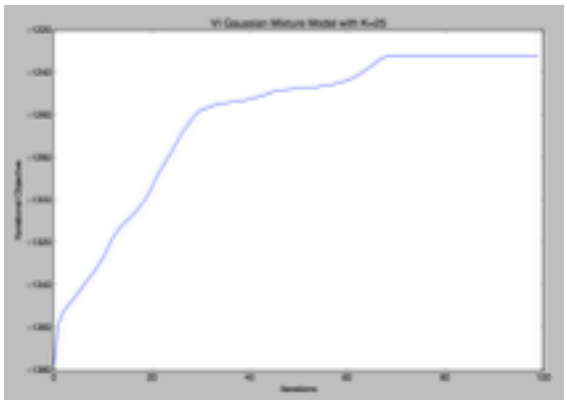
See code attached.

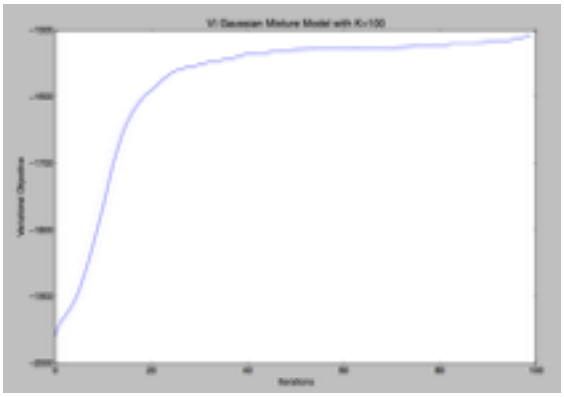
b) For each K, plot the variational objective function over the 100 iterations. What pattern do you observe?





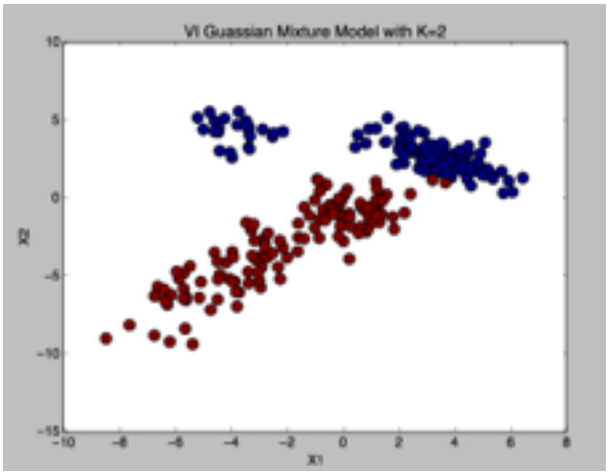


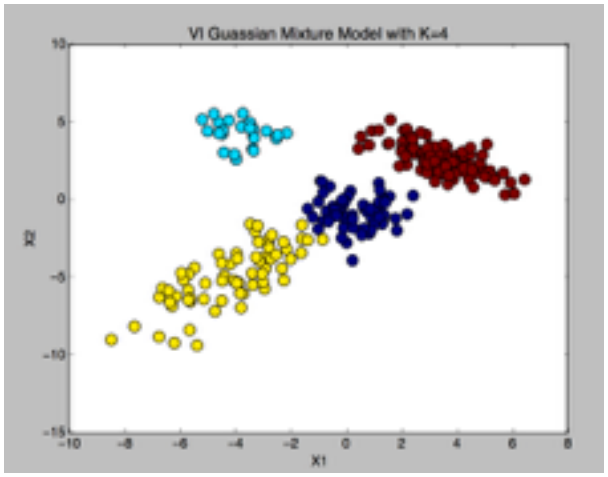


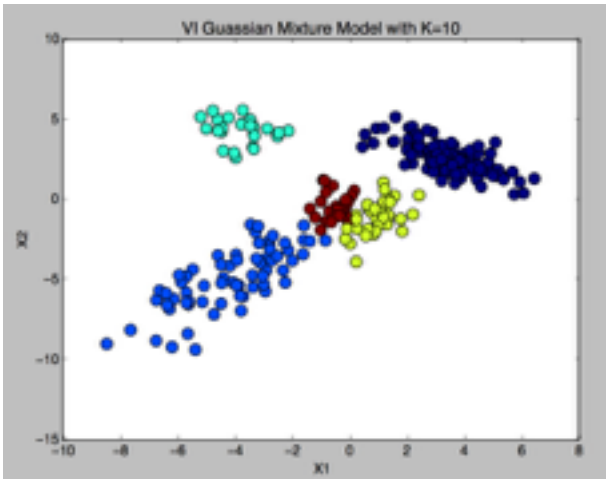


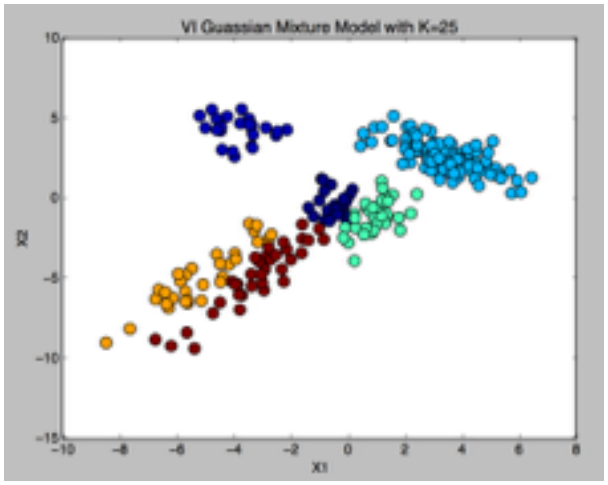
For all k's the log-likelihood increases.

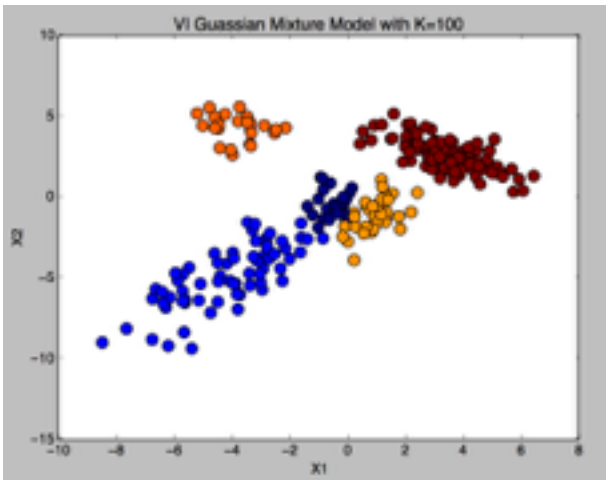
c) For the final iteration of each model, plot the data and indicate the most probable cluster of each observation according to $q(c_i)$ by a cluster-specific symbol. What do you notice about these plots as a function of K?











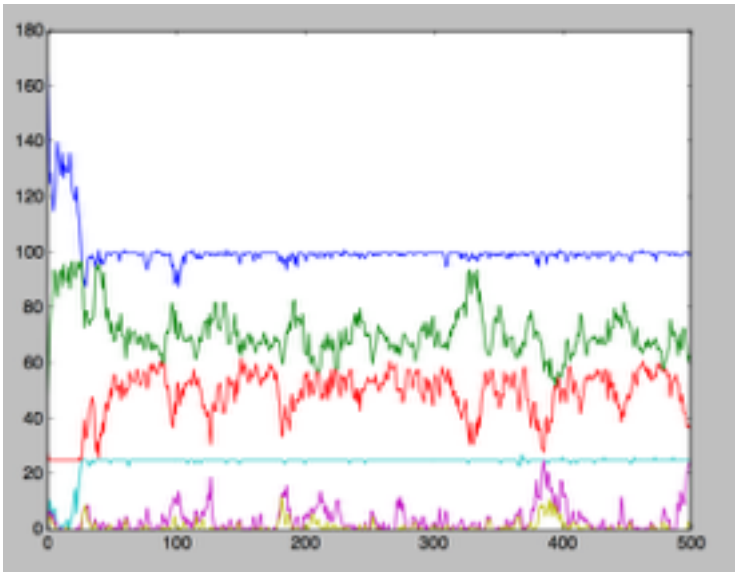
When plotting the clusters we can see that often times there are less than K clusters.

Problem 3:

a) Implement the above-mentioned Gibbs sampling algorithm discussed in class and described in the notes. Run your algorithm on the data provided for 500 iterations.

See code attached.

b) Plot the number of observations per cluster as a function of iteration for the six most probable clusters. These should be shown as lines that never cross; for example the ith value of the "second" line will be the number of observations in the second largest cluster after completing the ith iteration. If there are fewer than six clusters then set the remaining values to zero.



c) Plot of the total number of clusters that contain data as a function of iteration.