Bayesian Model for Machine Learning

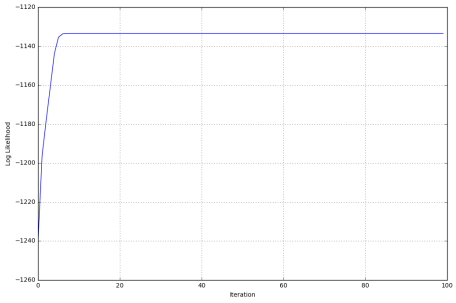
Haoyang Chen | hc2812 | HW4

1.

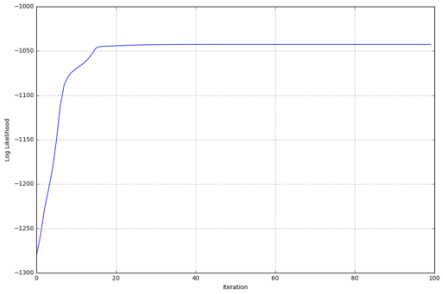
a). See attached code

b).

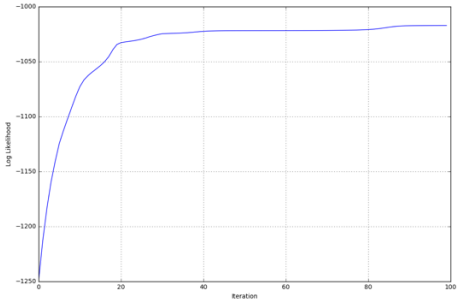
K = 2



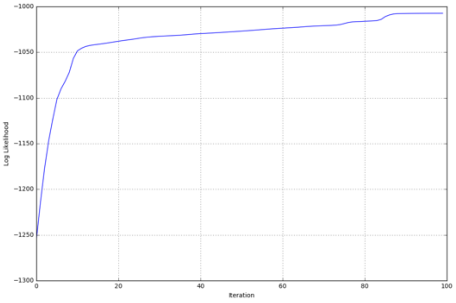
K = 4



K = 8



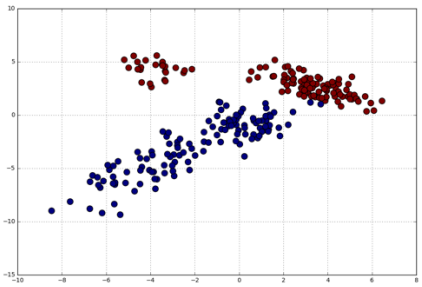
K = 10



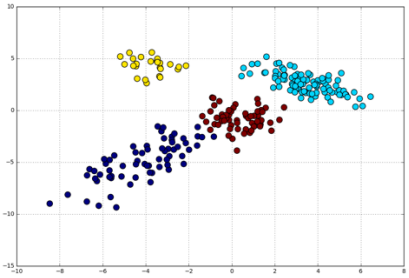
As K increases, the log-likelihood strictly increase and the converged rate would be slower. This would lead to overfitting to train data. It is not the best way to do model selection as if we use this method to choose model, we would always choose a larger K. Obviously, it is not a good idea to select K = n.

c).

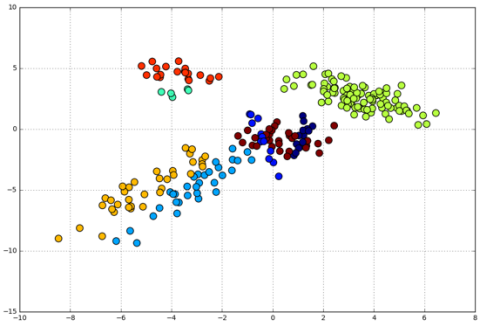
K = 2



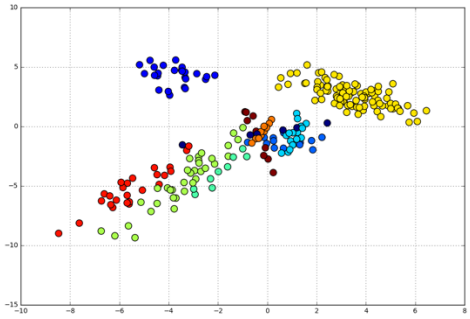
K = 4



K = 8



K = 10



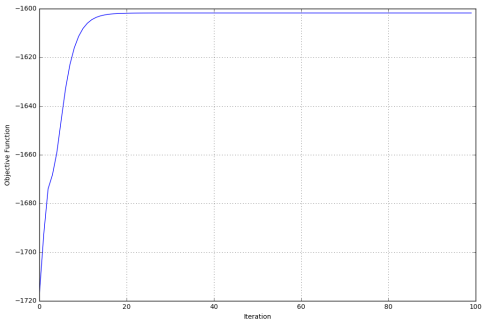
As K increases, the clusters tend to be smaller and concentrated but less intuitive. It would minimize overall error. As K increases, the clusters spread out over the data, but reflect less underlying structure.

2.

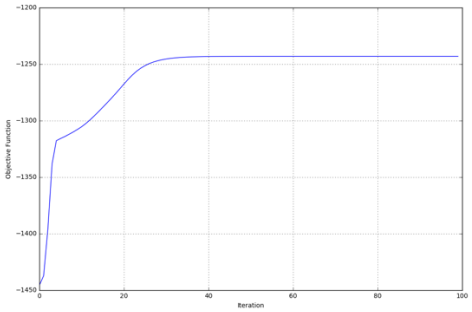
a). See attached code

b).

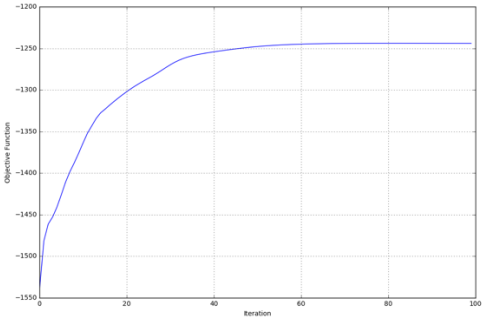
K = 2



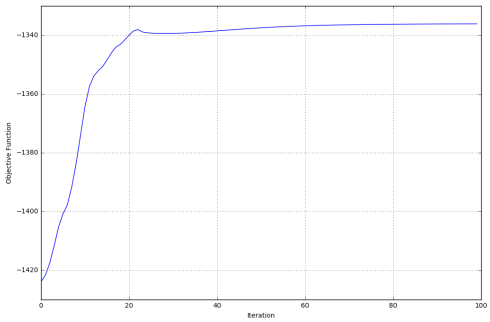
K = 4



K = 10



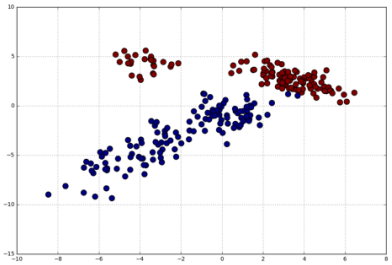
K = 25



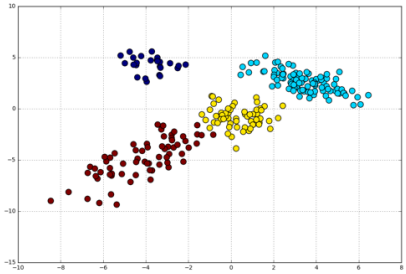
Unlike EM algorithm, the objective function is not monotonically increasing during the fitting process. Moreover, the converged objective function is not always increasing as K increases. It is hard to use objective function to select K.

c).

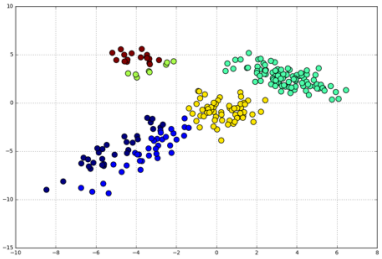
K = 2



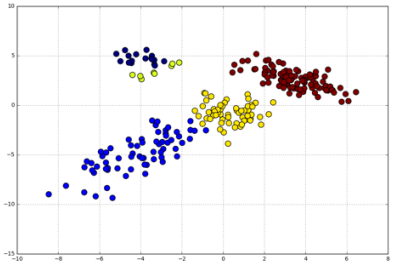
K = 4



K = 10



K = 25



Based on the plot, I think the number of clusters is not positively related to the number of K. When K = 25, the plot shows there are 5 clusters. However, when K = 10, still 5 clusters. The number of clusters is smaller than K. Generally, the algorithm cannot cluster too many clusters.

3.

a). See attached code.

b).

c).