ML Homework 3 Solution

Han Wang (hw2435@nyu.edu)

October 13, 2019

Problem 1

For the E-step

$$\begin{aligned}\tau_{n,j} &= p(z_n = j|x_n, \theta)\\ &= \frac{\pi_j p(x_n|\mu_j)}{\sum_i \pi_i p(x_n|\mu_j)}\\ &= \frac{\pi_j \prod_i \mu_j(l)^{x_n(l)}}{\sum_i \pi_i \prod_i \mu_j(l)^{x_n(l)}}\end{aligned} \tag{1-1}$$

For the M-step

$$\begin{aligned}\theta &= argmax_\theta \sum_n^N \sum_j^K \tau_{n,j} log \frac{p(x_n, z_n = j|\theta)}{\tau_{n,j}}\\ &= argmax_\theta \sum_n^N \sum_j^K \tau_{n,j} log \frac{p(x_n|z = j, \theta)p(z_n|\theta)}{\tau_{n,j}}\\ &= argmax_\theta \sum_n^N \sum_j^K \tau_{n,j} log \frac{\pi_j \mu_j^{x_n}}{\tau_{n,j}}\end{aligned} \tag{1-2}$$

Which is equivalent to

$$argmax_{\mu,\pi} \sum_n^N \sum_j^K \tau_{n,j}(log(\mu_j^{x_n}) + log(\pi_j)) \tag{1-3}$$

$$subject\ to\ \sum_{l}^{M} \mu_j(l) = 1,\ \sum_{i}^{K} \pi_i = 1. \tag{1-3*}$$

Write it with Lagrange multipliers:

$$L = \sum_{n}^{N} \sum_{j}^{K} \tau_{n,j}(log(\mu_j^{x_n}) + log(\pi_j)) - \alpha(\sum_{i}^{K} \pi_i - 1) - \sum_{j}^{K} \beta_j(\sum_{l}^{M} \mu_j(l) - 1) \tag{1-4}$$

Take the derivative along π_j:

$$\frac{\partial L}{\partial \pi_j} = 0$$

$$\sum_{n}^{N} \frac{\tau_{n,j}}{\pi_j} - \alpha = 0 \tag{1-5}$$

$$\pi_j = \frac{\sum_n^N \tau_{n,j}}{\alpha}$$

Plugged into the constraint:

$$\pi_j = \frac{1}{N} \sum_{n}^{N} \tau_{n,j} \tag{1-5*}$$

Take the derivative along $\mu_j(l)$

$$\frac{\partial L}{\partial \mu_j(l)} = 0$$

$$\sum_{n}^{N} \tau_{n,j} \frac{x_n(l)}{\mu_j(l)} - \beta_j = 0 \tag{1}$$

$$\mu_j(l) = \frac{1}{\beta_j} \sum_{n}^{N} \tau_{n,j} x_n(l)$$

Plugged into the constraint:

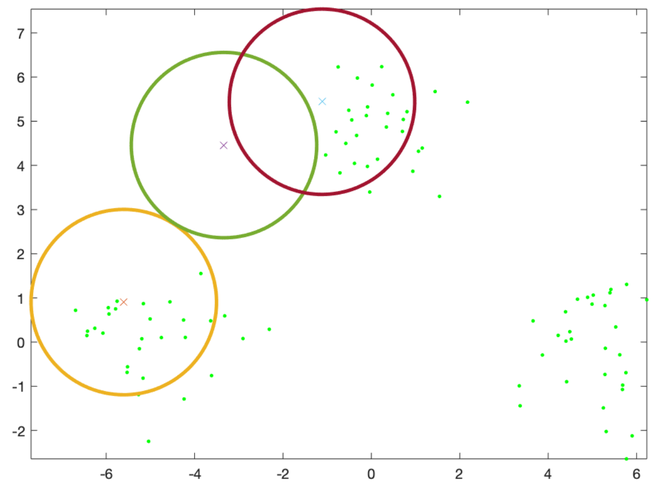
$$\mu_j(l) = \sum_{n}^{N} \frac{\tau_{n,j}}{\sum_n^N \tau'_{n,j}} x_n(l) \tag{2}$$

Hence, all the parameters are recovered.

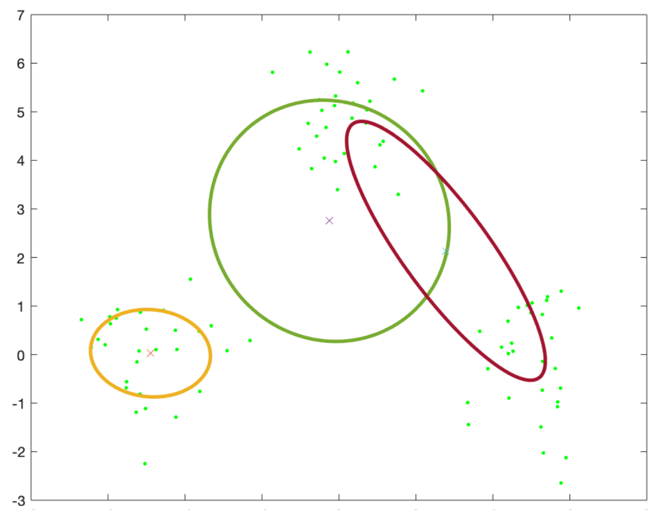
Problem 2

Part A

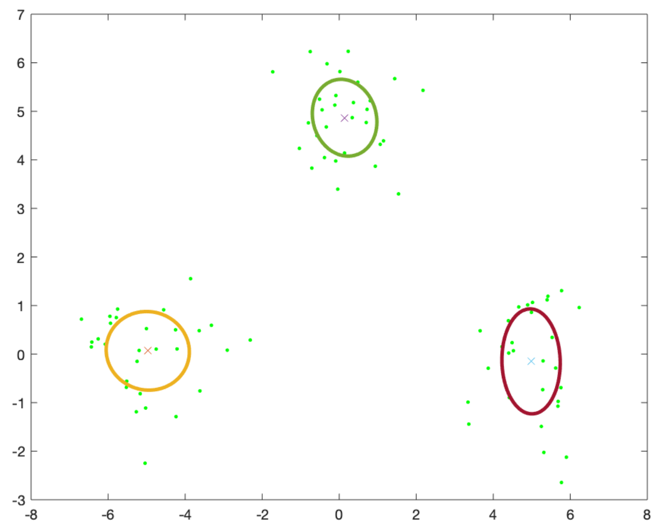
Dataset A - Init(Iteration=0)



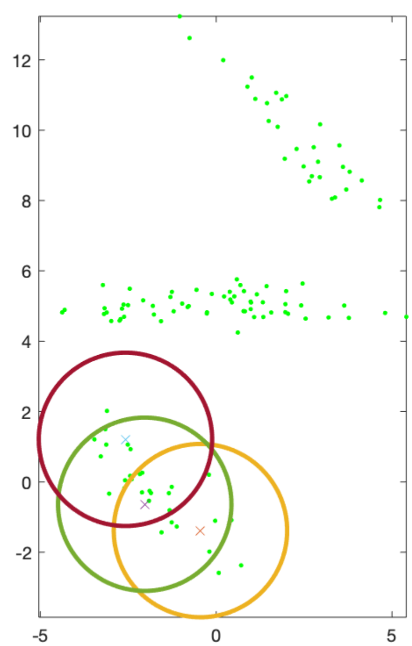
Dataset A - after one iteration



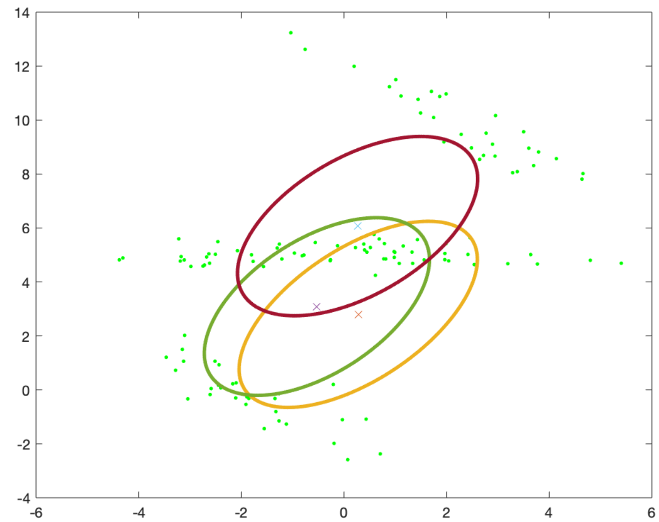
Dataset A - Converged



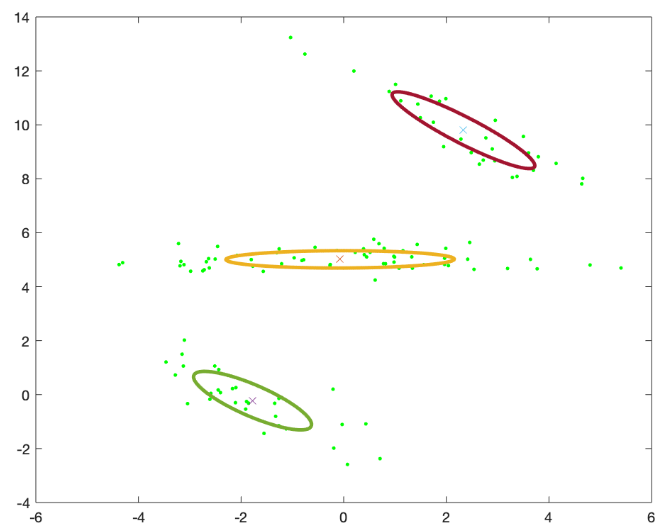
Dataset B - Init(Iteration=0)



Dataset B - after one iteration



Dataset B - Converged

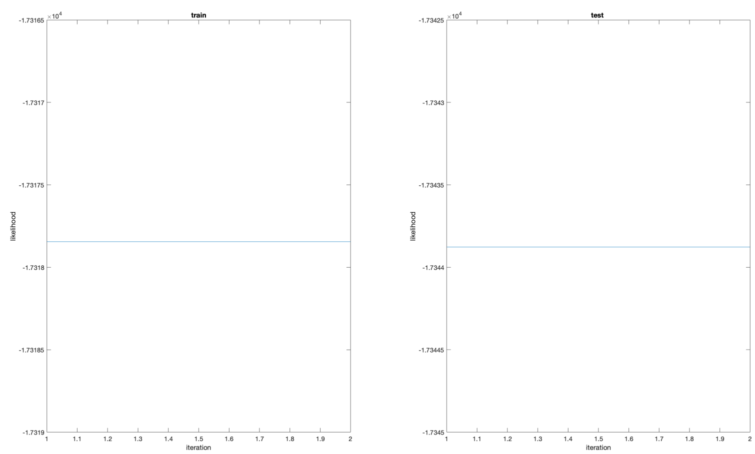


Part B

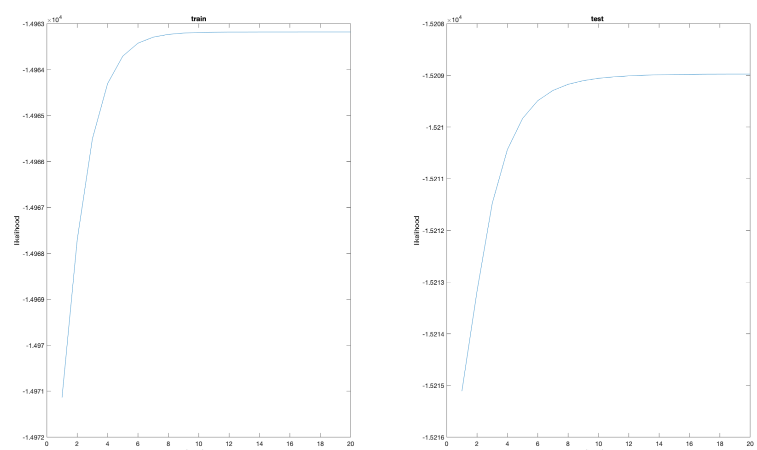
In this part, I modified the code to implement a new EM algorithm for clustering Bernoulli models.

For each experiment, I fix the value of K from $1, 2, 3, 4, 5$ to compute the training likelihood and testing likelihood.

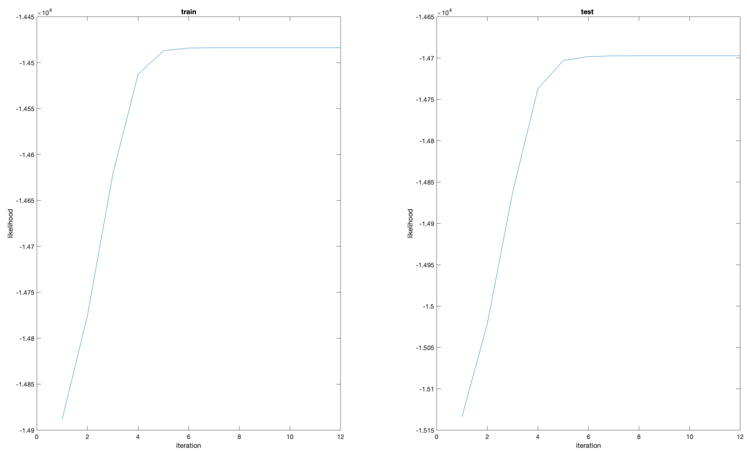
When $K = 1$, we get the plot with x-axis as iteration and y-axis as likelihood:



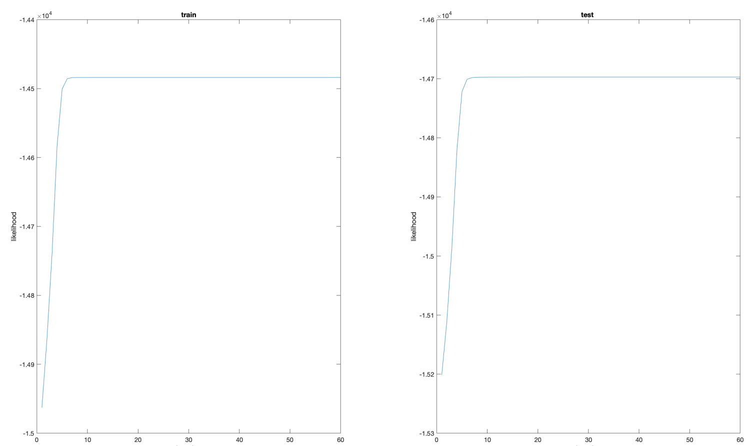
When $K = 2$, we get the plot with x-axis as iteration and y-axis as likelihood:



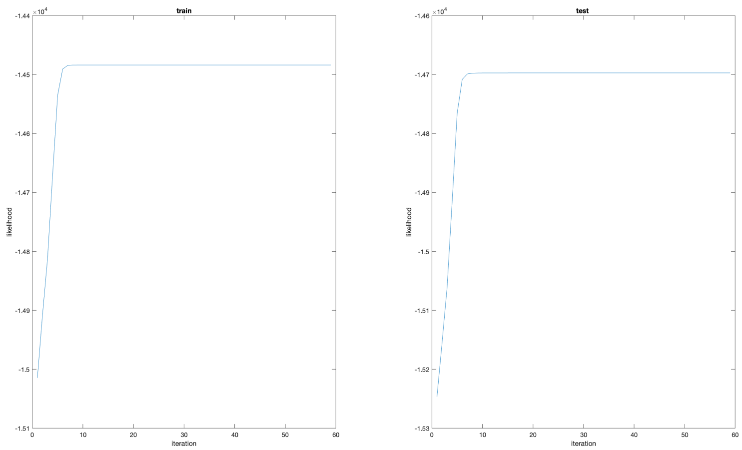
When $K = 3$, we get the plot with x-axis as iteration and y-axis as likelihood:



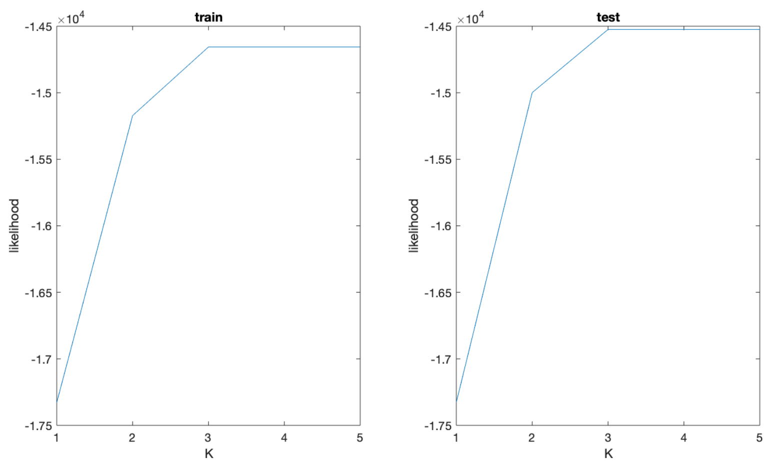
When $K = 4$, we get the plot with x-axis as iteration and y-axis as likelihood:

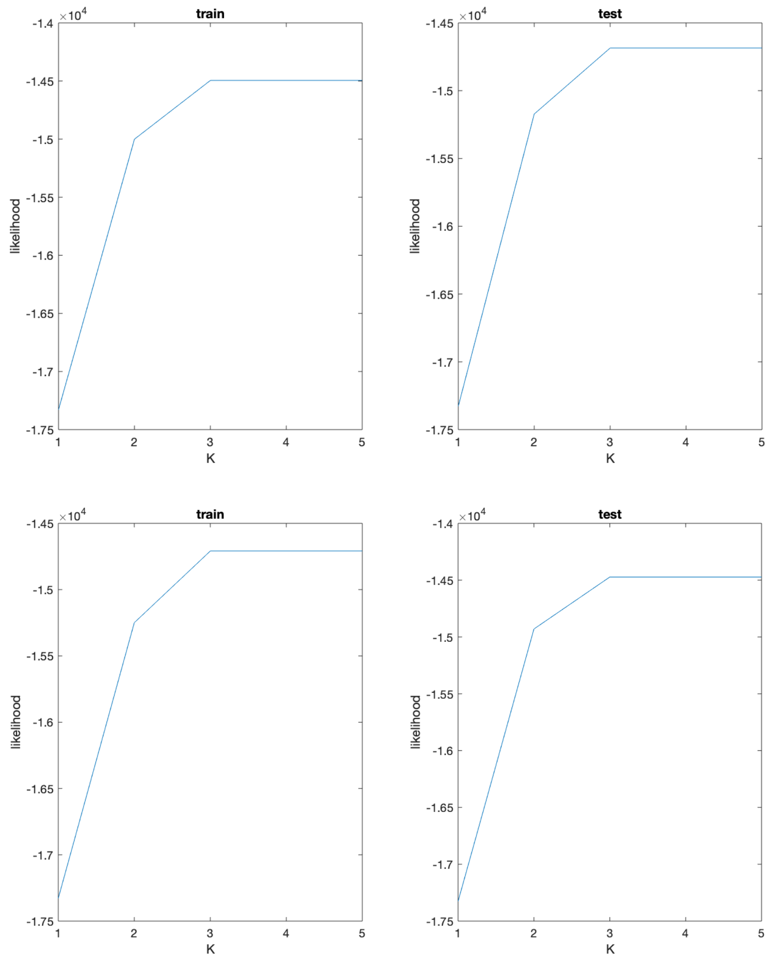


When $K = 5$, we get the plot with x-axis as iteration and y-axis as likelihood:



Also, we apply cross-validation to determine the best K with 3 various random initializations.





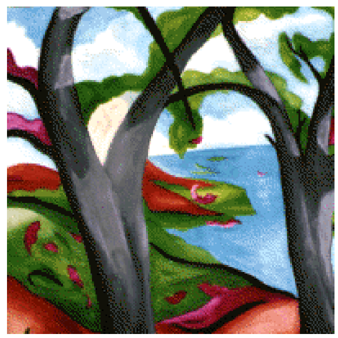
By the graph, we can tell that when K=3, the testing likelihood stopped increasing. So, the best K would probably be 3.

Problem 3

The data inconsistencies comes from that some of the initiated centers are never favored. Thus, when the center has been updated, there is no data points in the cluster, the new mean will be NaN.

The more smart way is to generate centers more seperately.

The original image is:



Apply K-Means with K=2:

Apply K-Means with K=4:

Apply K-means with K=6:

The way to improve the segmentation is to take the image locality into consideration, which is not only focus on the channel but also with the spatial features of the image.

Problem 4

a)

The arithmetic mean is $\frac{\sum x_i}{n}$, the geometric mean is $\sqrt[n]{\prod x_i}$, apply $log(f(x)) \geq f(log(x))$, which is a Jensen's inequality,

$$
\begin{aligned}
log(\frac{\sum x_i}{n}) &\geq \frac{\sum log(x_i)}{n} \\
log(\frac{\sum x_i}{n}) &\geq \frac{1}{n} log(\prod x_i) \\
log(\frac{\sum x_i}{n}) &\geq \log(\sqrt[n]{\prod x_i})
\end{aligned}
\tag{4-a}
$$

which is $\frac{\sum x_i}{n} \geq \sqrt[n]{\prod x_i}$.

b)

For $\alpha_i = \frac{exp(\theta^T f_i)}{\sum exp(\theta^T f_i)}$, we can knew $\sum \alpha_i = 1$. Take $f(x) = log(x)$, we get the Jensen's inequality as $f(E(x)) \geq E(f(x))$

$$
\begin{aligned}
\log(\sum exp(\theta^T f_i)) = \log(\sum \alpha_i \frac{exp(\theta^T f_i)}{\alpha_i}) &\geq \sum \alpha_i \log(\frac{exp(\theta^T f_i)}{\alpha_i}) \\
\log(\sum exp(\theta^T f_i)) &\geq \theta^T \sum \alpha_i f_i - \sum \alpha_i \log(\alpha_i) \\
\sum exp(\theta^T f_i)) &\geq exp(\theta^T \sum \alpha_i f_i - \sum \alpha_i \log(\alpha_i))
\end{aligned}
\tag{4-b}
$$