## 1. Problem 1

E-step:

$$\tau_{nj} = P(z_n = j \mid x_n, \theta) = \frac{P(x_n \mid z_n = j, \theta) \cdot P(z_n = j \mid \theta)}{P(x_n \mid \theta)} = \frac{\pi_j \, P(x_n \mid u_j)}{\sum_{i=1}^{K} \pi_i \, P(x_n \mid u_i)}$$

denote $U_j = \prod_{s=1}^{M} u_j(s)^{x_n(s)}$

so $\tau_{nj} = \dfrac{\pi_j U_j}{\sum_{i=1}^{K} \pi_i U_i}$

M-step:

$$P(x_n, z = j \mid \theta) = P(x_n \mid z = j, \theta) \cdot P(z_n = j \mid \theta) = \pi_j U_j$$

so $\theta := \arg\max_{\theta} \sum_{n=1}^{N} \sum_{j=1}^{K} \tau_{nj} \log \dfrac{P(x_n, z = j \mid \theta)}{\tau_{nj}}$

$:= \arg\max_{\theta} \sum_{n=1}^{N} \sum_{j=1}^{K} \tau_{nj} (\log \pi_j + \log U_j)$ s.t. $\sum_{s=1}^{M} u_j(s) = 1$, $\sum_{j=1}^{k} \pi_j = 1$

Applying Lagrangian:

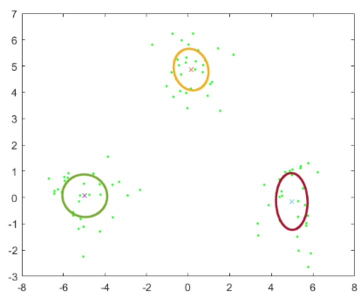
$$L(u, \pi, \alpha, \beta) = \sum_{n=1}^{N} \sum_{j=1}^{K} \tau_{nj} (\log \pi_j + \log U_j) - \alpha \left( \sum_{j=1}^{K} \pi_j - 1 \right) - \sum_{j=1}^{K} \beta_j \left( \sum_{s=1}^{M} u_j(s) - 1 \right)$$

$$\frac{\partial L}{\partial \pi_j} = 0 \Rightarrow \pi_j = \frac{1}{\alpha} \sum_{n=1}^{N} \tau_{nj} \Rightarrow \pi_j = \frac{1}{N} \sum_{n=1}^{N} \tau_{nj}$$

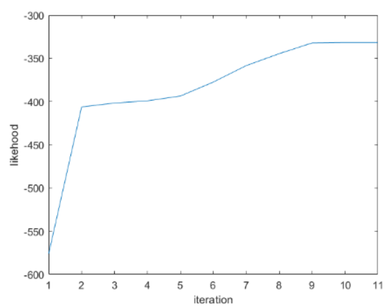
$$\frac{\partial L}{\partial u_j(s)} = 0 \Rightarrow u_j(s) = \frac{1}{\beta_j} \sum_{n=1}^{N} \tau_{nj} x_n(s) \Rightarrow u_j(s) = \sum_{n=1}^{N} \frac{x_{nj}}{\sum_{p=1}^{N} \tau_{pj}} x_n(s)$$

## 2. Problem 2

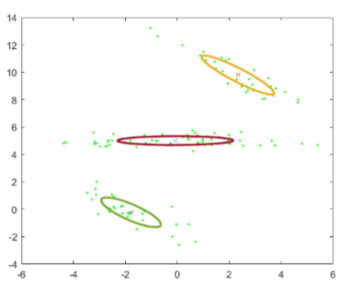
### Part A



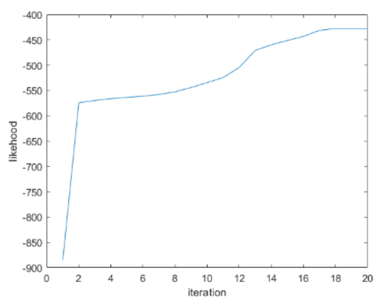
Dataset A




iteration result

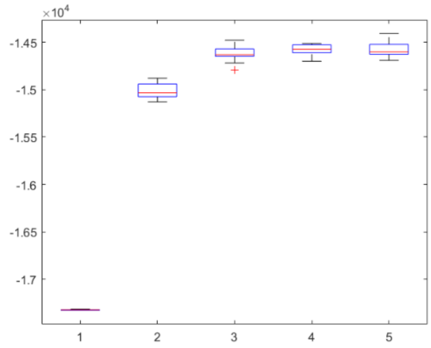


Dataset B

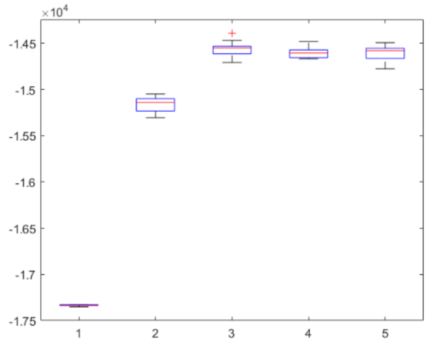



iteration result

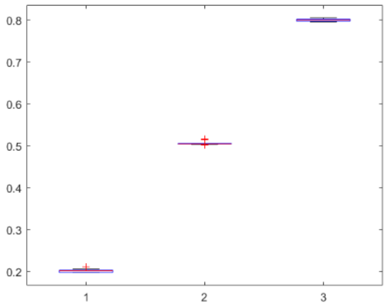
**Part B**




train log-likelihood average and std



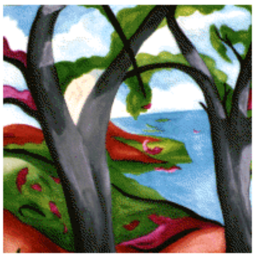

test log-likelihood average and std




The Bernoulli coefficients are [0.2, 0.5, 0.8] in average

Optimal coefficients average and std

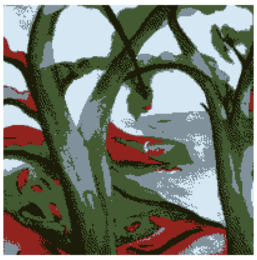
## 3. Problem 3



Original picture

k = 3

K = 4



k = 5

For randomly initialized K means, some values might be too far away from data points thus lead to it is not closest mean to any point, this explains the inconsistencies.

Too improve, we should calculate a range according to the given dataset for the initial means. Also, we should make sure that every initial mean will not be too close to each other.

4. **Problem 4**

(a)

We are proving $\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}$

applying Jensen's inequality

$$\log\left(\frac{\sum x_i}{n}\right) \geq \frac{1}{n}\sum \log(x_i) = \frac{1}{n}\log(\Pi x_i) = \log(\Pi x_i)^{1/n}$$

so proved.

(b)

$$\sum_{i=1}^{m} \exp(\theta^T f_i) = \sum_{i=1}^{m} \alpha_i \frac{\exp(\theta^T f_i)}{\alpha_i}, \quad \text{where } \sum \alpha_i = 1$$

So applying Jensen's inequality:

$$\ln\left[\sum_{i=1}^{m} \alpha_i \frac{\exp(\theta^T f_i)}{\alpha_i}\right] \geq \sum_{i=1}^{m} \alpha_i \ln\left[\frac{\exp(\theta^T f_i)}{\alpha_i}\right] = \theta^T \sum_{i=1}^{m} \alpha_i f_i - \sum_{i=1}^{m} \alpha_i \ln \alpha_i$$

$$= \ln\left[\exp\left(\theta^T \sum_{i=1}^{m} \alpha_i f_i - \sum_{i=1}^{m} \alpha_i \ln \alpha_i\right)\right]$$

$$\Rightarrow \sum_{i=1}^{m} \exp(\theta^T f_i) \geq \exp\left(\theta^T \sum_{i=1}^{m} \alpha_i f_i - \sum_{i=1}^{m} \alpha_i \ln \alpha_i\right)$$