Assignment 5

Name: Guangbo Yu

ID: 7262891072

1 K-means

1.1

$$\frac{\partial D}{\partial \mu_k} = \sum_{n=1}^{N}\sum_{k=1}^{K} -2r_{nk}(x_n - \mu_k) = 0$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}x_n - \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\mu_k = 0$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\mu_k = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}x_n$$

let k is an arbitrary value, we have

$$\sum_{n=1}^{N} r_{nk}\mu_k = \sum_{n=1}^{N} r_{nk}x_n$$

$$\mu_k = \frac{\sum_{n=1}^{N} r_{nk}x_k}{\sum_{n=1}^{N} r_{nk}}$$

which is the mean of all data points assign to the cluster k

1.2

$$\frac{\partial D}{\partial \mu_k} = \sum_{n=1}^{N} r_{nk}(I_{\mu_k \geq x_n} - I_{\mu_k \leq x_n}) = 0$$

$$\sum_{n=1}^{N} r_{nk}I_{\mu_k \geq x_n} = \sum_{n=1}^{N} r_{nk}I_{\mu_k \leq x_n}$$

which means the number of x less than μ_k is equal to x larger than μ_k, therefore μ_k is the median of the responding cluster

1.3

(a)

$$\tilde{D} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|\phi(x_n) - \tilde{\mu_k}\|_2^2$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}[\phi(x_n) - \tilde{\mu_k}]^T[\phi(x_n) - \tilde{\mu_k}]$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}[\phi(x_n) - \frac{\sum_{i=1}^{N}\phi(x_i)}{\sum_{n=1}^{N} r_{nk}}]^T[\phi(x_n) - \frac{\sum_{i=1}^{N}\phi(x_i)}{\sum_{n=1}^{N} r_{nk}}]$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\{\phi(x_n)^T\phi(x_n) - \frac{2\sum_{i=1}^{N}\phi(x_n)^T\phi(x_i)}{\sum_{n=1}^{N} r_{nk}} + \sum_{i=1}^{N}\frac{\phi(x_i)^T\phi(x_i)}{(\sum_{n=1}^{N} r_{nk})^2}\}$$

$$\tilde{D} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\{k(x_n, x_n) - \frac{2\sum_{i=1, i\in cluster_k}^{N} k(x_n, x_i)}{\sum_{n=1}^{N} r_{nk}} + \sum_{i=1 j=1 i,j\in cluster_k}^{N}\frac{k(x_i, x_j)}{(\sum_{n=1}^{N} r_{nk})^2}\}$$

(b) we will assign this point to closet centroid

$$k = argmin_k\|x - \mu_k\|_2^2$$

$$= argmin_k\|\phi(x) - \frac{\sum_{n=1}^{N} r_{nk}\phi(x_n)}{\sum_{n=1}^{N} r_{nk}}\|_2^2$$

$$= argmin_k\{\phi(x)^T\phi(x) - \frac{2\sum_{n=1}^{N} r_{nk}\phi(x_n)^T\phi(x)}{\sum_{n=1}^{N} r_{nk}} + \frac{\sum_{n=1}^{N} r_{nk}\phi(x_n)^T\phi(x_n)}{(\sum n = 1^N r_{nk})^2}\}$$

$$= argmin_k\{k(x, x) - \frac{2\sum_{n=1}^{N} r_{nk}k(x_n, x)}{\sum_{n=1}^{N} r_{nk}} + \frac{\sum_{n=1}^{N} r_{nk}k(x_n, x_n)}{(\sum n = 1^N r_{nk})^2}\}$$

(c) Input:
K: kernel matrix
$C_i : iclusters$
k: number of clusters

Code:
1. random initialize partition points$\{x_1, \ldots, x_n\}$ into k clusters
2. while not converged:
3. For all points $x_n, = 1, 2, \ldots, n$:
4. For all clusters C_i, $i = 1, \ldots, k$:
5. compute distance: $\|\phi(x_n) - \tilde{\mu_i}\|_2^2$ using result of (a)
6. end for
7. $C(x_n) = argmin_k\|\phi(x_n) - \mu_k\|_2^2$ using result of (b)
8. end for
9.return clusters

2 GMM

$$L(x_1) = \alpha\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + (1-\alpha)\frac{1}{\sqrt{\frac{\pi}{2}}}e^{-2x^2}$$

$$\alpha = p(z_n = k|x_n) = \frac{p(x_n|z_n = k)p(z_n = k)}{\sum p(x_n|z_n = k')p(z_n = k')}$$

$$= \frac{p(\theta_1)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}{p(\theta_1)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + p(\theta_2)\frac{1}{\sqrt{\frac{\pi}{2}}}e^{-2x^2}}$$

3 EM

(a)
let $z_i = 1$ when zero state, $z_i = 1$ when imperfect state

so likelihood for zero state and Poisson distribution is:

$$L_1(\pi, \lambda; x, Z) = \prod_{x_i=0} \{z_i\pi + (1-z_i)(1-\pi)e^{-\lambda}\}$$

$$L_2(\pi, \lambda; x, Z) = \prod_{x_i>0} \{(1-z_i)(1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\}$$

so likelihood for Zero-inflated Poisson distribution is:

$$L(\pi, \lambda; x, Z) = \prod_{x_i=0} \{z_i\pi + (1-z_i)(1-\pi)e^{-\lambda}\} + \prod_{x_i>0} \{(1-z_i)(1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\}$$

and the log likelihood is:

$$l(\pi, \lambda; x, Z) = \sum_{i=0}^{n} \log(z_i\pi) + \sum_{i=0}^{n} \log(1-z_i) + \sum_{i=0}^{n} \log(1-\pi) + \sum_{i=0}^{n} \{x_i\lambda - \lambda - \log(x_i!)\}$$

(b)
E-step:

$$p(z_i = 1|x_i; \pi, \lambda) = \frac{P(x_i|zero)P(zero)}{P(x_i|zero)P(zero) + P(x_i|Poisson)P(Poisson)}$$

$$\gamma_{ik} = P(z_i|x_i; \pi, \lambda) = \frac{\pi}{\pi + (1-\pi)e^{-\lambda}}$$

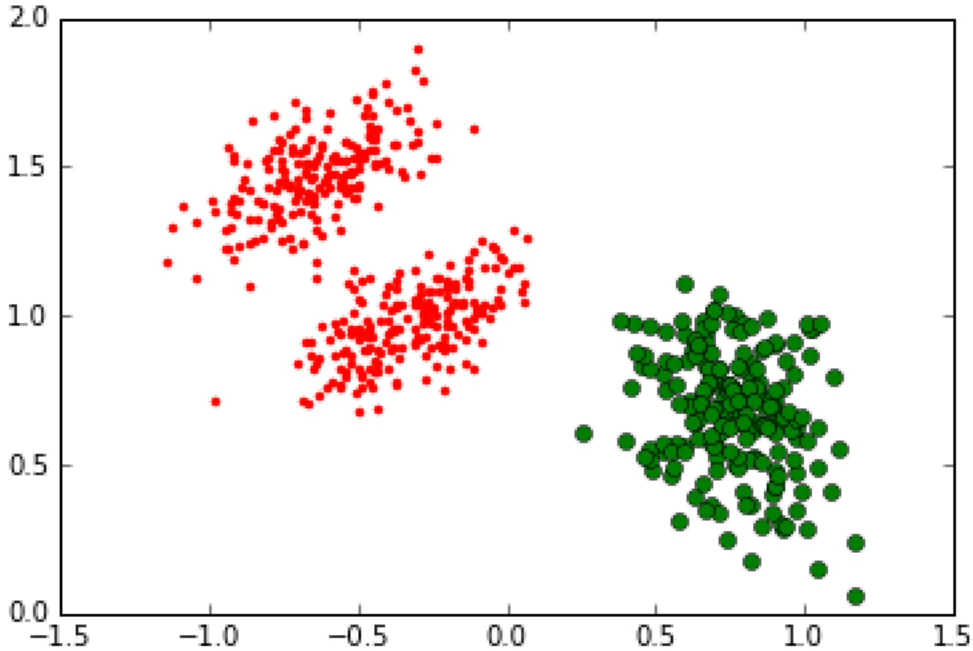
M-step:

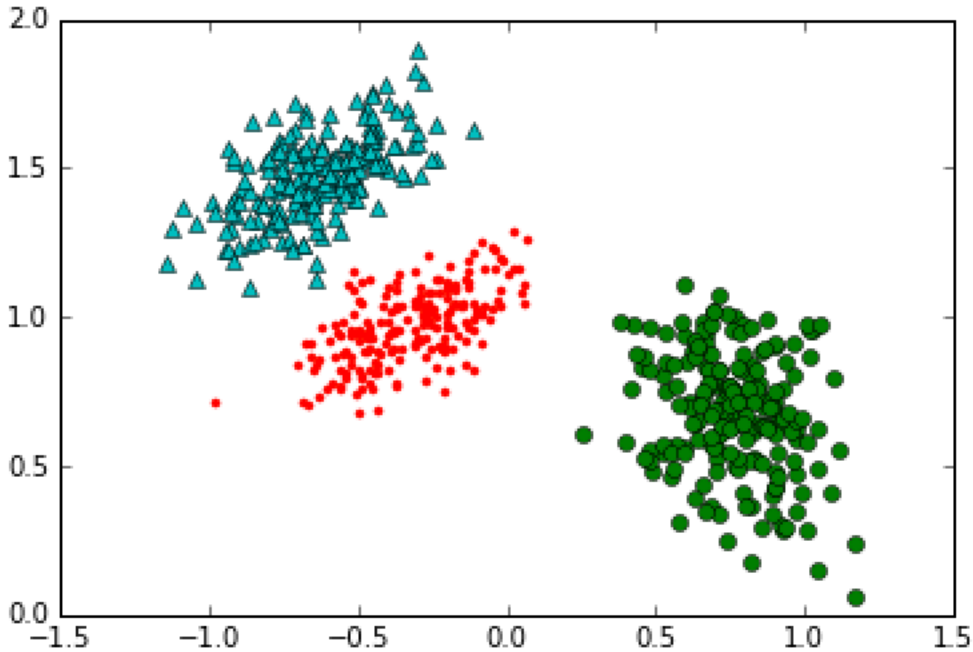
$$\pi, \lambda = argmax_{\pi,\lambda} Q(\theta, \theta^{old}) = argmax_{\pi,\lambda} \sum_{n}\sum_{k} \gamma_{ik} \log((\frac{\pi}{\pi + (1-\pi)e^{-\lambda}})^{x_k}(\frac{(1-\pi)e^{-\lambda}}{\pi + (1-\pi)e^{\lambda}})^{1-x_k})$$

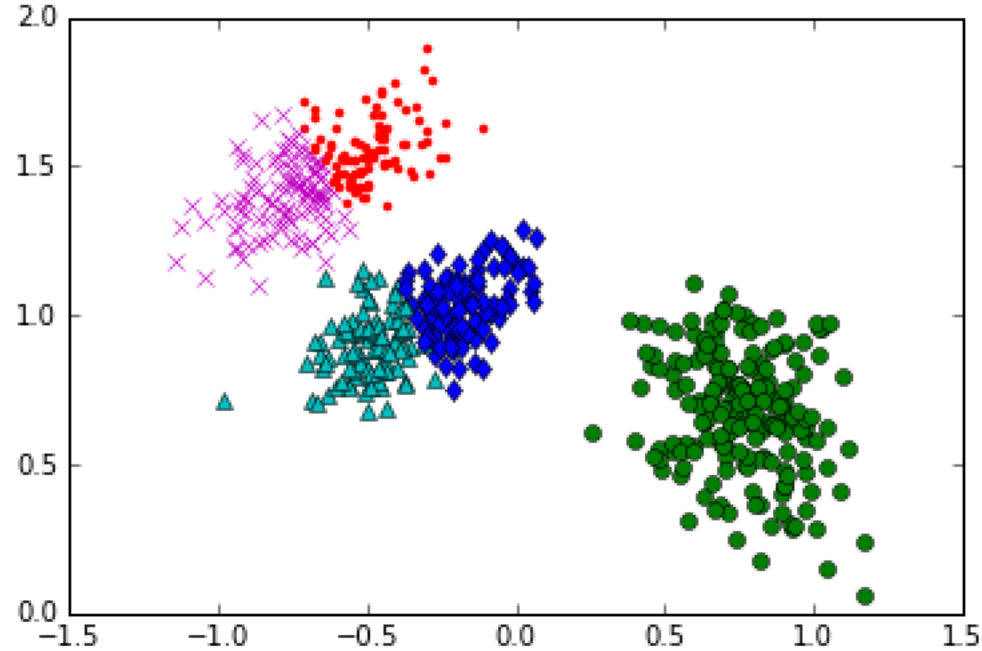
Programming

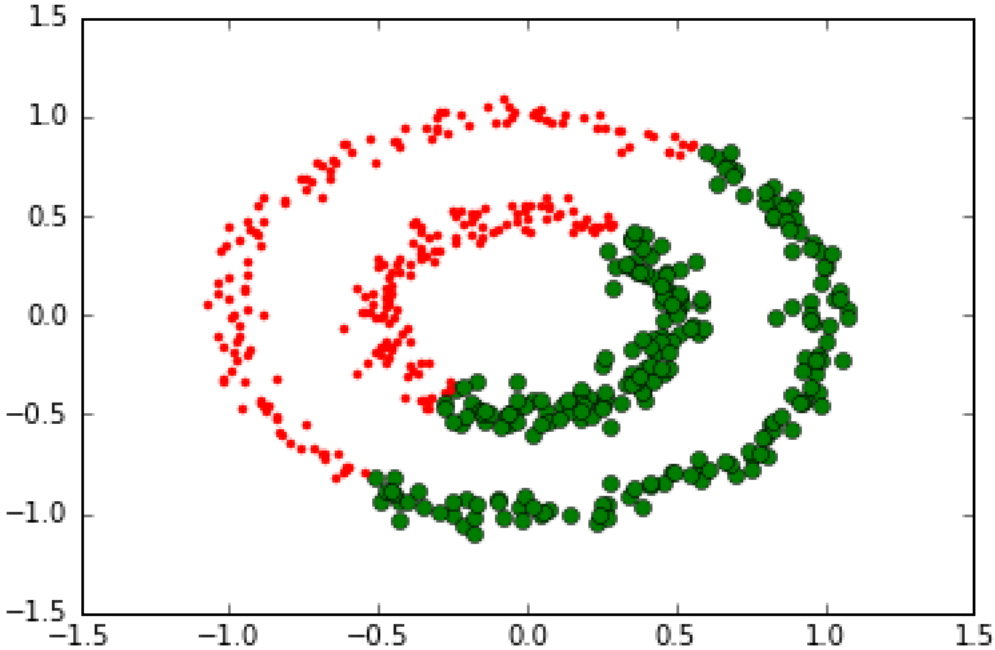
4.2

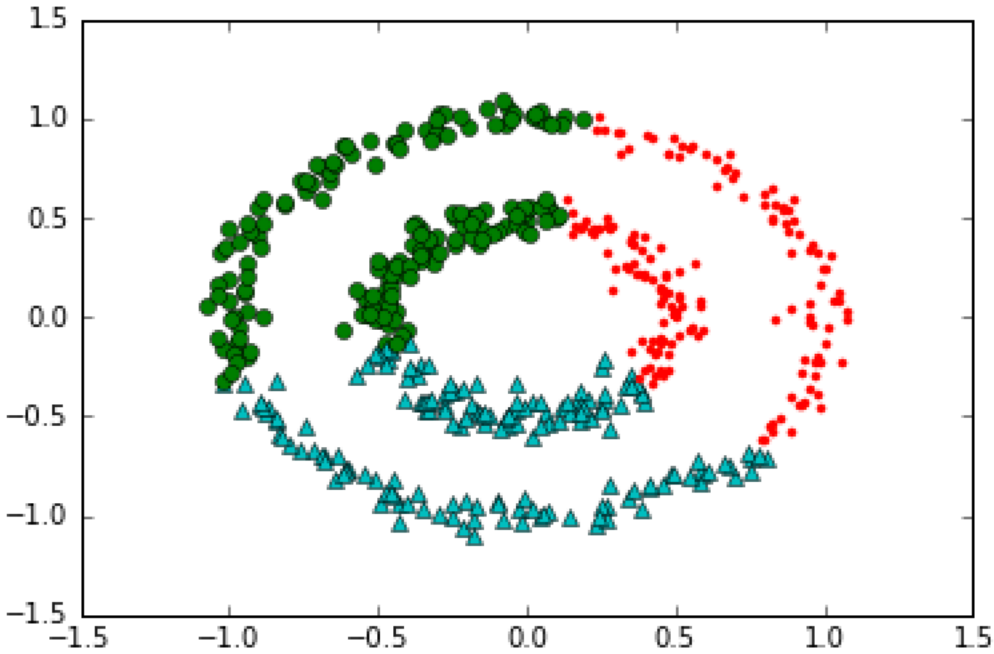
(a)

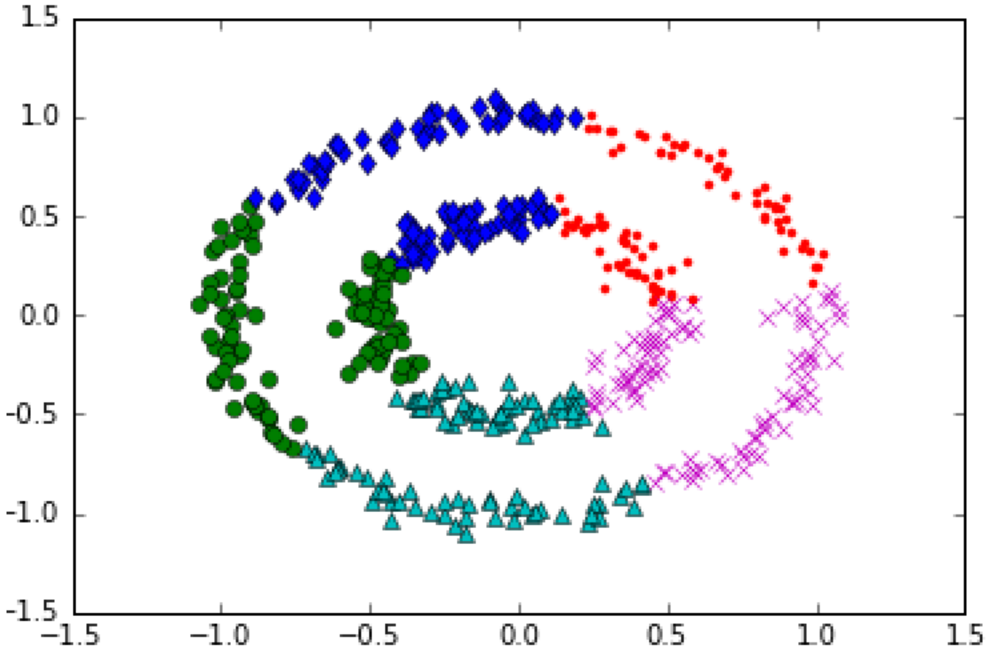





2.0
1.5
1.0
0.5
0.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5





(b)

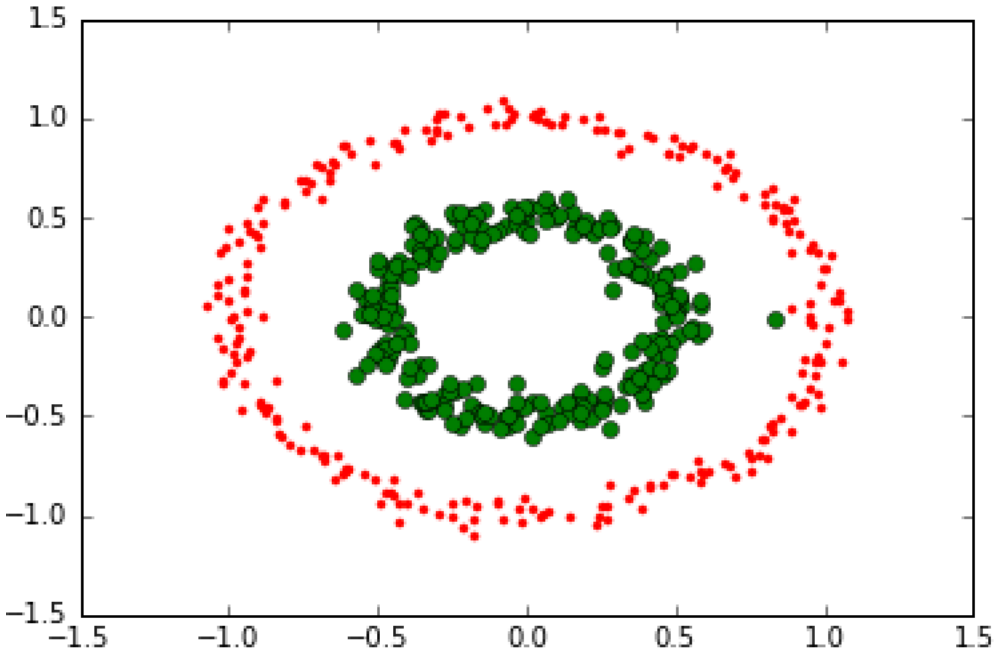
because hw_circle.csv is not linear seperable, we cannot use Euclidean distance to cluster it

4.3

(a)

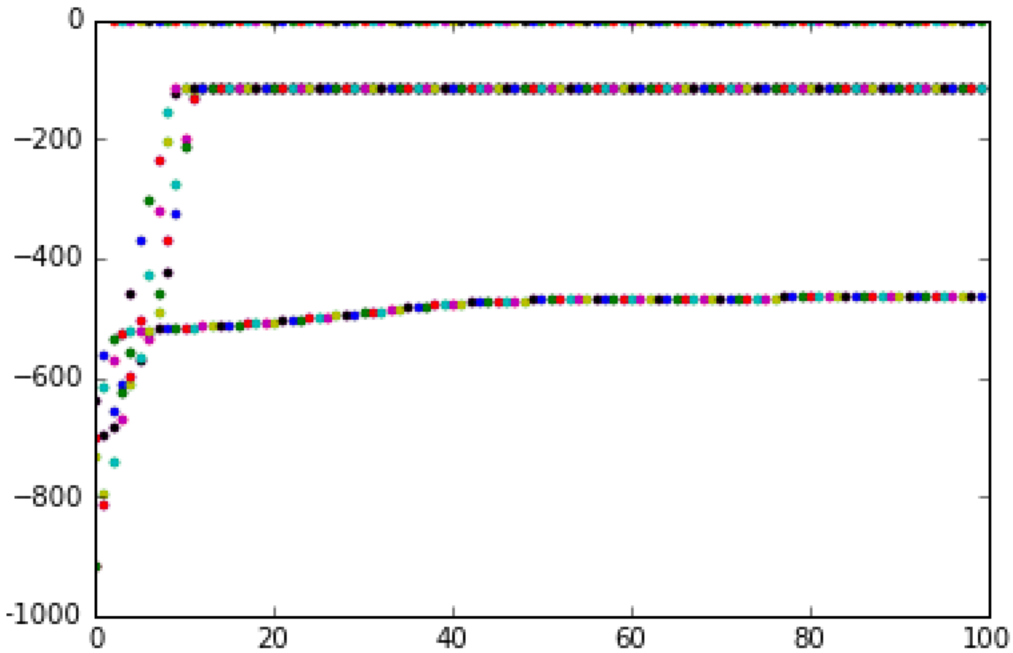
RBF kernel

(b)



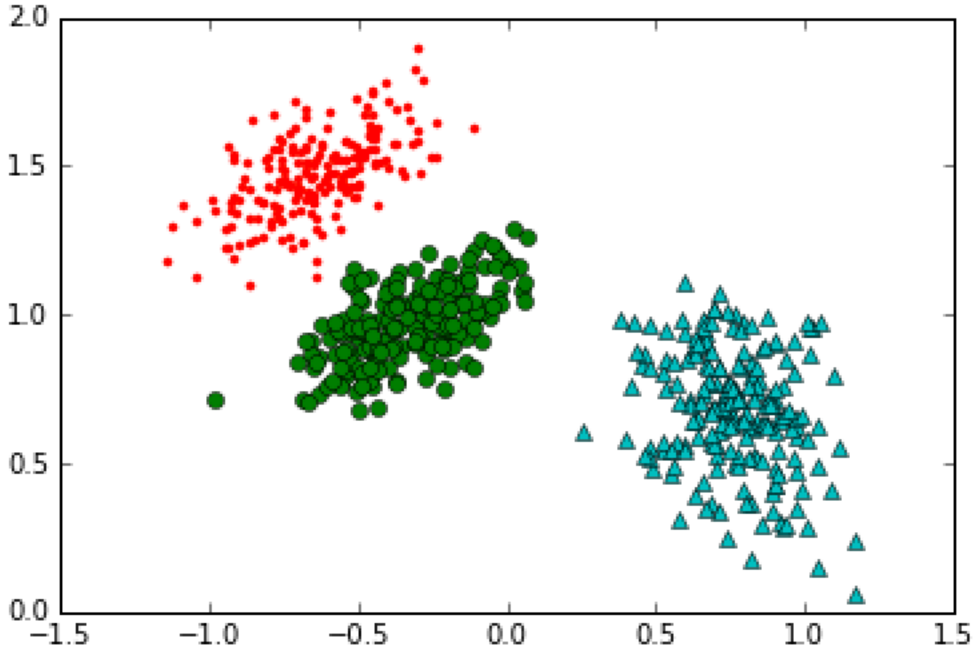
4.4

(a)



max log likelihood is around -115

(b)



Means :

[0.75896032,   0.67976982],

[-0.6394629 ,   1.4746064],

[-0.32592106,   0.97133574]

Covariance matrix:

$$\begin{matrix} 0.02717056 & -0.00840045 \\ -0.00840045 & 0.040442 \end{matrix}$$

$$\begin{matrix} 0.0359676 & 0.01549315 \\ 0.01549315 & 0.01935168 \end{matrix}$$

$$\begin{matrix} 0.03604954 & 0.01463887 \\ 0.01463887 & 0.0162912 \end{matrix}$$