a) $P(Z=1;\theta) = w \qquad P(Z=2;\theta) = 1-w$

joint likelihood

$$P(X=x, Z=1;\theta) = P(Z=1;\theta)\cdot P(X=x|Z=k;\theta) = w\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-u_1)^2}{2}}$$

$$P(X=x, Z=2;\theta) = P(Z=2;\theta)\,P(X=x|Z=k;\theta) = (1-w)\,\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-u_2)^2}{2}}$$

marginal likelihood

$$P(X=x;\theta) = \sum_{k=1}^{2} P(X=x, Z=k;\theta) = \sum_{k=1}^{2} P(Z=k;\theta)\,P(X=x|Z=k;\theta)$$

$$= w\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-u_1)^2}{2}} + (1-w)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-u_2)^2}{2}}$$

log-likelihood

$$\ell(X=x;\theta) = \log\sum_{k=1}^{2} P(X=x, Z=k;\theta)$$

n i.i.d.

$$\ell(X_1=x_1,\cdots,X_n=x_n;\theta) = \log\left(\prod_{i=1}^{n}\sum_{k=1}^{2} P(X_i=x_i, Z=k;\theta)\right) = \sum_{i=1}^{n}\log\sum_{k=1}^{2} P(X_i=x_i, Z=k;\theta)$$

(b) $\ell(X_i;\theta) = \log P(X=x_i;\theta) = \log \sum_{k=1}^{2} P(X=x_i, Z=k;\theta)$ let q_i be some distribution

$$= \log \sum_{k=1}^{2} \frac{P(X=x, k;\theta)}{q_i(k)} \cdot q_i(k)$$

$\sum_{k=1}^{2} q_i(k) = 1$, $q_i(k) \geq 0$

$$= \log \mathbb{E} \frac{P(X=x, k;\theta)}{q_i(k)}$$

Since log is concave, by Jensen's inequality

$\mathbb{E}\phi(X) \leq \phi\mathbb{E}(X)$

$$\geq \mathbb{E} \log \frac{P(X=x, k;\theta)}{q_i(k)} = F_i(\theta; q_i)$$

$$= \sum_{k=1}^{2} q_i(k) \cdot \log \frac{P(X=x, Z;\theta)}{q_i(k)}$$

$$= \sum_{k=1}^{2} q_i(k) \log P(X_i, k;\theta) + \sum_{k=1}^{2} q_i(k) \log\left(\frac{1}{q_i(k)}\right)$$

$$\ell(\{X_i\}_{i=1}^{n};\theta) \geq F(\theta; q) = \sum_{i=1}^{n} F_i(\theta; q_i)$$

c). for above Jensen's inequality to hold with equality, we want the expectation to be taken over a constant, such that equality holds

for a single observation x_i, let $\frac{P(X=x_i, Z=z_i; \theta)}{q_i(Z=z_i)} = C$

$$q_i(Z=z_i) \propto P(X=x_i, Z=z_i; \theta)$$

Since $q_i(Z=z_i)$ is a distribution, its summation is 1

$$q_i(Z=z_i) = \frac{P(X=x_i, Z=z_i; \theta)}{\sum_{i=1}^{n} P(X=x_i, Z=z_i; \theta)}$$

$$= \frac{P(X=x_i, Z=z_i; \theta)}{P(X=x_i; \theta)}$$

$$= P(Z=z_i \mid X=x_i; \theta)$$

Then for n i.i.d. sample, given θ^t

$$q^{t+1}(Z_1=z_1, \cdots, Z_n=z_n) = \prod_{i=1}^{n} P(Z=z_i \mid X=x_i; \theta^t)$$

(d) for given θ^t, we can calculate $P(X=x_i;\theta^t)$

$P(Z=k; X=x_i; \theta^t)$

$$P(Z=1|X=x_i;\theta^t) = \frac{P(Z=1, X=x_i;\theta^t)}{P(X=x_i;\theta^t)} = \frac{P(Z=1, X=x_i;\theta^t)}{\sum_{k=1}^{2} P(X=x_i, Z=k;\theta^t)}$$

$$= \frac{w^t \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u_1^t)^2}{2}}}{w^t \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u_1^t)^2}{2}} + (1-w^t)\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u_2^t)^2}{2}}}$$

$$P(Z=2|X=x_i;\theta^t) = \frac{(1-w^t)\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u_2^t)^2}{2}}}{w^t \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u_1^t)^2}{2}} + (1-w^t)\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u_2^t)^2}{2}}}$$

e)

$$\sum_{i=1}^{n} \ell(X_i; \theta, q_i^{t+1}) = L(\theta; q^{t+1}) = \sum_{i=1}^{n}\sum_{k=1}^{2} q_i(k) \log p(X_i, k; \theta)$$

$$= \sum_{i=1}^{n}\left[q_i^{t+1} \cdot \log\left(w \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(X-u_1)^2}{2}}\right) + (1-q_i^{t+1})\left(\log (1-w) \frac{1}{\sqrt{2\pi}} e^{-\frac{(X-u_2)^2}{2}}\right)\right]$$

$$= \sum_{i=1}^{n}\left[q_i^{t+1}\left(\log w - \log(\sqrt{2\pi}) - \frac{(X-u_1)^2}{2}\right) + (1-q_i^{t+1})\left(\log(1-w) - \log(\sqrt{2\pi}) - \frac{(X-u_2)^2}{2}\right)\right]$$

$$= \sum_{i=1}^{n}\Big[q_i^{t+1}\left(\log w - \frac{(X-u_1)^2}{2}\right) - q_i^{t+1}\log(\sqrt{2\pi}) + q_i^{t+1}\log(\sqrt{2\pi}) - \log(\sqrt{2\pi})$$

$$+ (1-q_i^{t+1})\left(\log(1-w) - \frac{(X-u_2)^2}{2}\right)\Big]$$

$$= \sum_{i=1}^{n}\left[q_i^{t+1}\left(\log w - \frac{(X-u_1)^2}{2}\right) + (1-q_i^{t+1})\left(\log(1-w) - \frac{(X-u_2)^2}{2}\right) - \log(\sqrt{2\pi})\right]$$

$$= C + \sum_{i=1}^{n}\left[q_i^{t+1}\left(\log w - \frac{(X-u_1)^2}{2}\right) + (1-q_i^{t+1})\left(\log(1-w) - \frac{(X-u_2)^2}{2}\right)\right]$$

(f)

$$\nabla_{u_1} L(\theta; q^{t+1}) = \nabla_{u_1} \sum_{i=1}^{n} q_i^{t+1}\left(\log w - \frac{(X_i-u_1)^2}{2}\right)$$

$$= \nabla_{u_1} \sum_{i=1}^{n} -q_i^{t+1} \frac{(X_i-u_1)^2}{2}$$

$$= \sum_{i=1}^{n} q_i^{t+1} \cdot (X_i - u_1) = 0$$

$$\sum_{i=1}^{n} q_i^{t+1} X_i = \sum_{i=1}^{n} q_i^{t+1} u_1$$

$$u_1^{t+1} = \frac{\sum_{i=1}^{n} q_i^{t+1} X_i}{\sum_{i=1}^{n} q_i^{t+1}}$$

$$\nabla_{u_2} L(\theta; q^{t+1}) = \nabla_{u_2} \sum_{i=1}^{n} (1-q_i^{t+1})\left(\log(1-w) - \frac{(X_i-u_2)^2}{2}\right)$$

$$= \nabla_{u_2} \sum_{i=1}^{n} -(1-q_i^{t+1}) \frac{(X_i-u_2)^2}{2}$$

$$= \sum_{i=1}^{n} (1-q_i^{t+1})(X_i-u_2) = 0$$

$$\sum_{i=1}^{n} (1-q_i^{t+1}) X_i = \sum_{i=1}^{n} (1-q_i^{t+1}) u_2$$

$$u_2^{t+1} = \frac{\sum_{i=1}^{n} (1-q_i^{t+1}) X_i}{\sum_{i=1}^{n} (1-q_i^{t+1})}$$

$$\nabla_{w} L(\theta; q^{t+1}) = \sum_{i=1}^{n}\left[q_i^{t+1} \cdot \frac{1}{w} + (1-q_i^{t+1}) \cdot \frac{-1}{(1-w)}\right] = 0$$

$$\frac{1}{w} \sum_{i=1}^{n} q_i^{t+1} = \frac{1}{1-w} \sum_{i=1}^{n} (1-q_i^{t+1})$$

$$\frac{1-w}{w} \sum_{i=1}^{n} q_i^{t+1} = n - \sum_{i=1}^{n} q_i^{t+1}$$

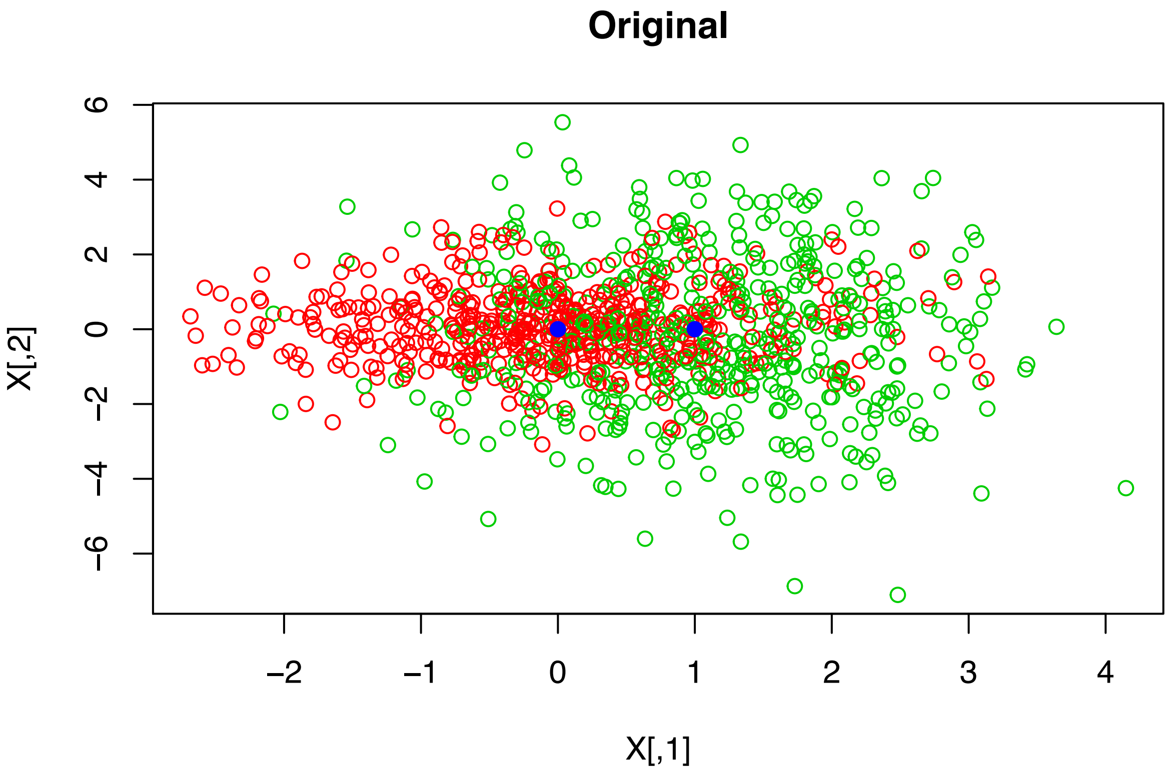
$$\frac{1}{w} - 1 = \frac{n}{\sum_{i=1}^{n} q_i^{t+1}} - 1$$

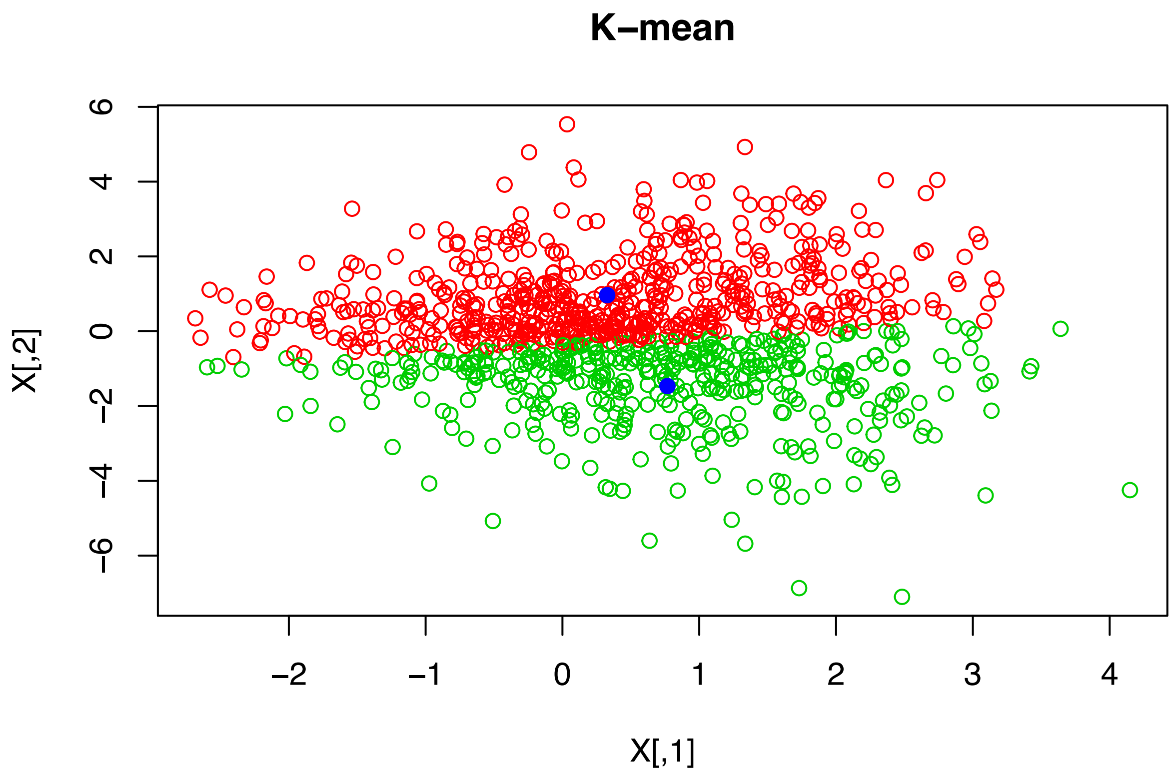
$$w^{t+1} = \frac{\sum_{i=1}^{n} q_i^{t+1}}{n}$$

problem(g)

caojilin

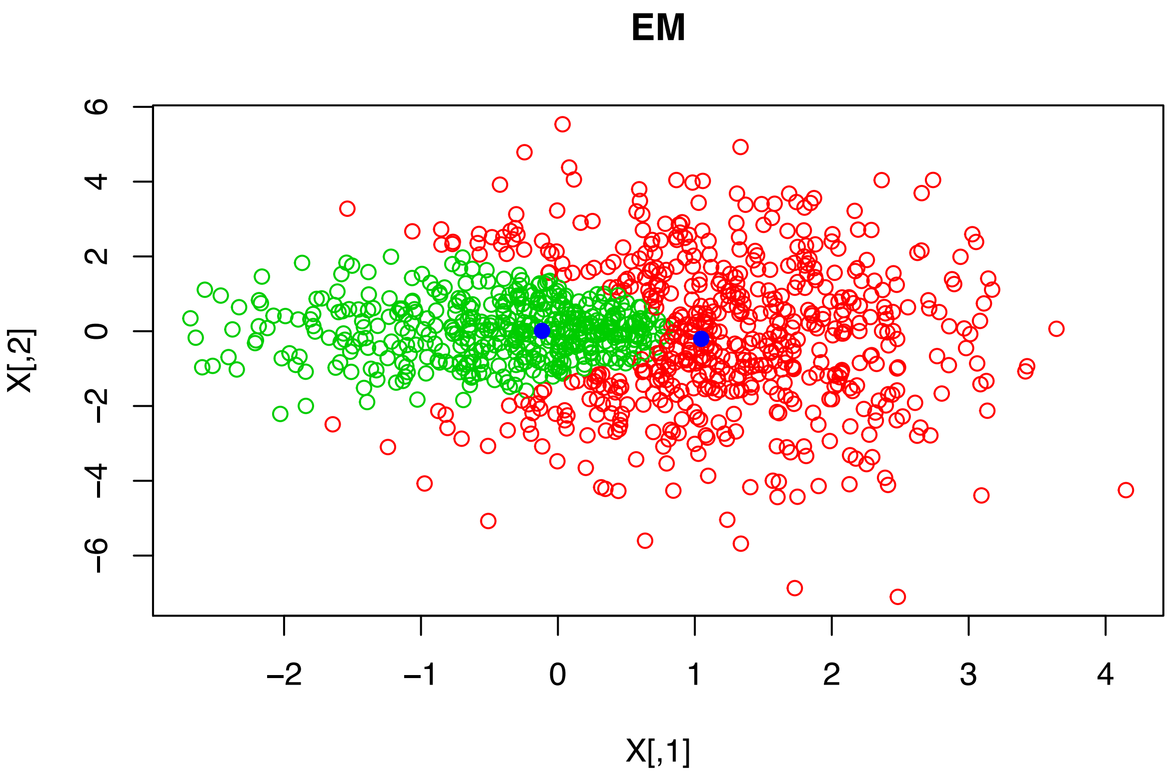
3/4/2019






```
## [1] "EM centers"

##            [,1]         [,2]
## [1,]  1.0467685 -0.203373117
## [2,] -0.1157559  0.008598925
```



K-means only calculates Euclidean distance, which is like a non-parametric approach, while EM assumes there exist underlying distributions and uses likelihood to calculate, which is more like a parametric approach. We know that original data is almost non-separable. K-mean gives us a “hard”" classification while EM gives us a “soft” classification. Thus the estimated labels are different.