$$\gamma(z_i^t) = \frac{\pi_i \cdot \frac{1}{2b_i} \cdot \exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)}{\sum_{j=1}^{K} \pi_j \cdot \frac{1}{2b_j} \cdot \exp\left(\frac{-|x^t - \mu_j|}{b_j}\right)}$$

1 $\quad p(x^t \mid C_i) = f(x^t \mid \mu_i, b_i) = \frac{1}{2b_i} \exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)$

$$p(x^t) = \sum_{i=1}^{K} p(x^t \mid C_i)\, p(C_i)$$

$$= \sum_{i=1}^{K} \pi_i \frac{1}{2b_i} \exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)$$

$$\mathcal{L}(\{\mu_i, b_i, \pi_i\}_{i=1}^{K} \mid x) = \sum_{t=1}^{N} \log\left[\sum_{i=1}^{K} \pi_i \frac{1}{2b_i} \exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)\right]$$

$\underline{\mu_i:}$

$$\frac{\partial \mathcal{L}}{\partial \mu_i} = \sum_{t=1}^{N}\left[\frac{\pi_i \cdot \frac{1}{2b_i} \exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)}{\sum_{j=1}^{K} \pi_j \frac{1}{2b_j} \exp\left(\frac{-|x^t - \mu_j|}{b_j}\right)} \cdot \frac{\partial}{\partial \mu_i}\left(\frac{-|x^t - \mu_i|}{b_i}\right)\right]$$

$$= \sum_{t=1}^{N}\left[\gamma(z_i^t) \frac{\partial}{\partial \mu_i}\left(\frac{-|x^t - \mu_i|}{b_i}\right)\right]$$

$$= \sum_{t=1}^{N}\left[-\frac{\gamma(z_i^t)}{b_i} \frac{\partial}{\partial \mu_i}\left(|x^t - \mu_i|\right)\right]$$

Now, let $S_i = \{ t \in \{1, \ldots, N\} \mid i = \arg\max_j \gamma(z_j^t) \}$

and $\beta_i^t = \begin{cases} 1 & \text{if } i = \arg\max_j \gamma(z_j^t) \\ 0 & \text{ow} \end{cases}$

Thus $\beta_i^t = \begin{cases} 1 & \text{if } t \in S_i \\ 0 & \text{ow} \end{cases}$

Continuing,

$$\frac{\partial \mathcal{L}}{\partial \mu_i} = \sum_{t=1}^{N} \left[\frac{-\gamma(z_i^t)}{b_i} \frac{\partial}{\partial \mu_i} (|x^t - \mu_i|) \right]$$

$$= \sum_{t \in S_i} \left[\frac{-1}{b_i} \frac{\partial}{\partial \mu_i} (|x^t - \mu_i|) \right]$$

$$= \frac{-1}{b_i} \left[\sum_{t \in S_i} \frac{\partial}{\partial \mu_i} (|x^t - \mu_i|) \right]$$

$$= \frac{-1}{b_i} \frac{\partial}{\partial \mu_i} \left[\sum_{t \in S_i} (|x^t - \mu_i|) \right] = 0$$

$$\frac{\partial}{\partial \mu_i} \left[\sum_{t \in S_i} (|x^t - \mu_i|) \right] = 0$$

This is the same as minimizing

$\sum_{t \in S_i} |x^t - \mu_i|$, which happens

when $\boxed{\mu_i = \underset{t \in S_i}{\mathrm{median}} (x^t)}$

b_i:

$$\frac{\partial \mathcal{L}}{\partial b_i} = \left[\sum_{t=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \cdot 1/2b_j \cdot \exp\left(-\frac{|x^t-\mu_j|}{b_j}\right)} \cdot \frac{\pi_i}{2} \cdot \left(\frac{\partial (b_i^{-1})}{\partial b_i} \exp\left(-\frac{|x^t-\mu_i|}{b_i}\right) + \frac{1}{b_i}\frac{\partial}{\partial b_i}\left(\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right)\right)\right)\right]$$

$$\left[\frac{\partial (b_i^{-1})}{\partial b_i} \exp\left(\frac{-|x^t-\mu_i|}{b_i}\right) = \frac{-1}{b_i^2}\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right) = \frac{-1}{b_i}\left(\frac{1}{b_i}\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right)\right)\right.$$

$$\left[\frac{1}{b_i}\frac{\partial}{\partial b_i}\left(\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right)\right) = \frac{1}{b_i}\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right)\left(-|x^t-\mu_i| \cdot -\frac{1}{b_i^2}\right)\right.$$

$$= \frac{|x^t-\mu_i|}{b_i^2}\left(\frac{1}{b_i}\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right)\right)$$

$$\frac{\partial \mathcal{L}}{\partial b_i} = \sum_{t=1}^{N}\left[\frac{\pi_i \frac{1}{2b_i}\exp\left(-\frac{|x^t-\mu_i|}{b_i}\right)}{\sum_{j=1}^{K}\pi_j \cdot \frac{1}{2b_j}\exp\left(-\frac{|x^t-\mu_j|}{b_j}\right)} \cdot \left(\frac{|x^t-\mu_i|}{b_i^2} - \frac{1}{b_i}\right)\right]$$

$$= \sum_{t=1}^{N} \gamma(z_i^t)\left(\frac{|x^t-\mu_i|}{b_i^2} - \frac{1}{b_i}\right) = 0$$

$$\sum_{t=1}^{N} \gamma(z_i^t)\left(\frac{|x^t-\mu_i|}{b_i^2}\right) = \sum_{t=1}^{N}\gamma(z_i^t)\frac{1}{b_i}$$

$$\frac{1}{b_i^2}\sum_{t=1}^{N}\gamma(z_i^t)(|x^t-\mu_i|) = \frac{1}{b_i}\sum_{t=1}^{N}\gamma(z_i^t).$$

$$b_i = \frac{\sum_{t=1}^{N} \gamma(z_i^t)\left(|x^t - \mu_i|\right)}{\sum_{t=1}^{N} \gamma(z_i^t)}$$

π_i:

$$\frac{\partial}{\partial \pi_i}\left(\mathcal{L} - \alpha\left[\left(\sum_{i=1}^{k} \pi_i\right) - 1\right] \right) =$$

$$\sum_{t=1}^{N}\left[\frac{\left(\frac{1}{2b_i}\right)\exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)}{\sum_{j=1}^{k} \pi_j \frac{1}{2b_j} \exp\left(\frac{-|x^t - \mu_j|}{b_j}\right)}\right] - \alpha = 0$$

$$\sum_{t=1}^{N}\left[\frac{\pi_i \frac{1}{2b_i} \exp\left(\frac{-|x^t - \mu_i|}{b_i}\right)}{\sum_{j=1}^{k} \pi_j \frac{1}{2b_j} \exp\left(\frac{-|x^t - \mu_j|}{b_j}\right)}\right] = \pi_i \alpha$$

$$\sum_{t=1}^{N} \gamma(z_i^t) = \pi_i \alpha$$

$$\pi_i = \frac{1}{\alpha} \sum_{t=1}^{N} \gamma(z_i^t)$$

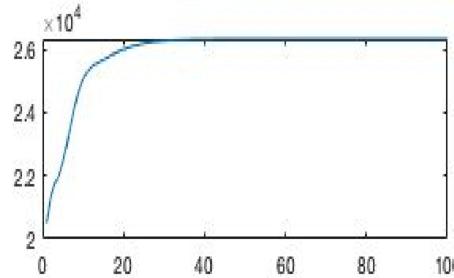
$$1 = \sum_{i=1}^{k} \pi_i = \frac{1}{\alpha} \sum_{i=1}^{k} \sum_{t=1}^{N} \gamma(z_i^t) = N/\alpha = 1$$

$$\alpha = N$$

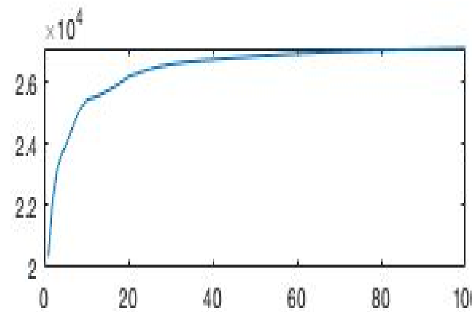
$$\boxed{\pi_i = \frac{\sum_{t=1}^{N} \gamma(z_i^t)}{N}}$$

2a)

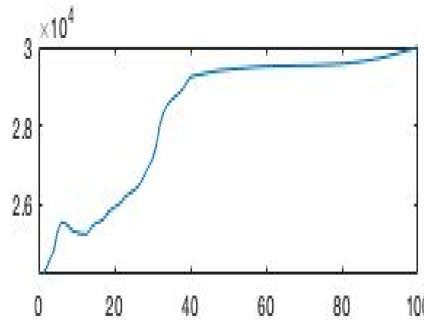
K = 4


×10^4
2.6
2.4
2.2
2
0
20
40
60
80
100

K = 8

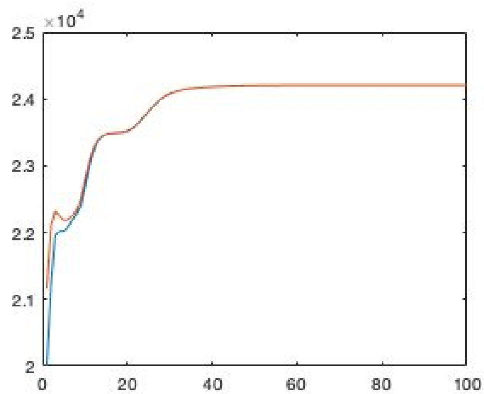

×10^4
2.6
2.4
2.2
2
0
20
40
60
80
100

K = 12

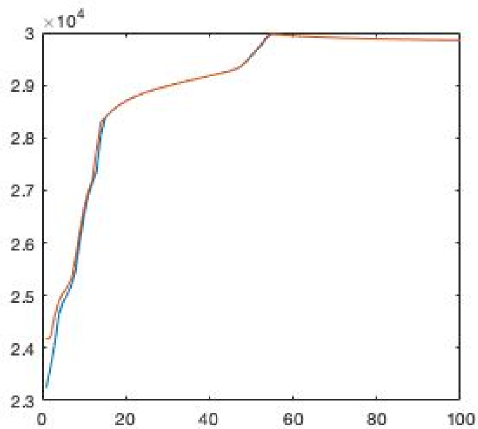

×10^4
3
2.8
2.6
0
20
40
60
80
100

2b) In these graphs, the blue line corresponds to the values of the log likelihood after the E step, and the new orange line corresponds to the values of the log likelihood after the M step. For the first few iterations, each orange value is noticeably higher than the blue value for the same iteration. This is as expected, since the parameters are updated in the M step, so the value of the log likelihood is higher after updating the parameters than it was before. However, after the first few iterations, the values converge very closely. With every iteration, the marginal increase of the log likelihood function gets smaller and smaller, so the difference between the log likelihood values before and after the parameter update gets smaller as well.

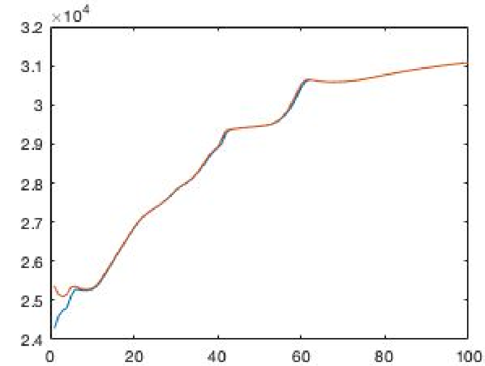
K = 4



K = 8



K = 12



2c) For this image, the initial learned covariance matrix is singular. In the EM algorithm, the inverse of this matrix is used in the calculation of h_i^t , and an uninvertable matrix throws an error. In k-means, we are not using soft indicator variables, so our b_i^t's depend only on the means and the training data, and the uninvertable matrix is not used anywhere in the algorithm. Thus, EM throws an error, while k-means runs successfully and produces the following reconstructed image.

2d $\mathcal{L}(\mu, \Sigma | x) =$

$$\sum_{t=1}^{N} \Big[-\frac{d}{2} \log 2\pi - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (x^t - \mu)^T \Sigma^{-1} (x^t - \mu) - \frac{\lambda}{2} \sum_{i=1}^{k} \sum_{j=1}^{d} (\Sigma_{ij}^{-1}$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \sum_{t=1}^{N} \left[\frac{1}{2} \Sigma - \frac{1}{2} (x^t - \mu)(x^t - \mu)^T - \frac{\lambda I}{2} \right] = 0$$

$$\frac{N}{2} \Sigma - \frac{1}{2} \sum_{t=1}^{N} (x^t - \mu)(x^t - \mu)^T - \frac{N \lambda I}{2} = 0$$

$$N \Sigma = N \lambda I + \sum_{t=1}^{N} (x^t - \mu)(x^t - \mu)^T$$

$$\boxed{\Sigma = \lambda I + \frac{1}{N} \sum_{t=1}^{N} (x^t - \mu)(x^t - \mu)^T}$$

2e)
After adding the regularizing term, the covariance matrix is no longer singular, so the algorithm is able to complete without raising any errors. I chose lambda to be .001, so it did not have any noticeable impact on the resulting image when compared to k-means. When I chose lambda to be too high, the image was not learned properly and the reconstructed image was a solid cream colored block. When I chose lambda to be too small, I still encountered error messages saying that the covariance matrix was uninvertable. The image below was produced using modified EM with lambda = .001.