Question 1:

Probability density function of a Laplacian distribution for class C_i (given):

$$p(x|C_i) = f(x|\mu_i, b_i) = \frac{1}{2b_i} \exp\left(-\frac{|x-\mu_i|}{b_i}\right), \quad b_i > 0$$

Mixture density of K Laplacian distributions (given):

$$P(x) = \sum_{i=1}^{k} \pi_i \frac{1}{2b_i} \exp(-\frac{|x-\mu_i|}{b_i}), \quad \Sigma_{i=1}^{k} \pi_i = 1$$

$$\boldsymbol{X} = \{x^1, x^2, \ldots, x^t, \ldots, x^N\}$$

$$P(X = \{x^1, \ldots x^N\}|C_i) = \frac{m!}{x^1! \cdots x^N!} = p_{i1}^{x^1} \cdots p_{in}^{x^N}$$

$$P(X) = \sum_{i=1}^{k} p(x|C_i) P(C_i)$$

Log-likelihood:

$$L(\phi|x) = \Sigma_{\mathrm{t}} \log p(x^t|\phi)$$
$$= \Sigma_t \log \sum_{i=1}^{k} \pi_i \frac{m! p_{i1}^{x^t} \cdots p_n^{x^t}}{x_1^t! \cdots x_n^t!}$$

$$\frac{\partial L}{\partial p_{ij}} = \Sigma_t \frac{\pi_i \frac{m! p_{i1}^{x^t} \cdots p_n^{x^t}}{x_1^t! \cdots x_n^t!}}{\Sigma_{i=1}^{k} \frac{\pi_i m! p_{i1} x^t \cdots p_n^{x^t}}{x_1^t \cdots x_n^t!}} + \alpha = 0$$

$$\Longrightarrow 0 = \Sigma_t \gamma(2_i^t) \frac{x_j^t}{p_{ij}} + \alpha$$
$$= \Sigma_{\mathrm{j}} \Sigma_t \gamma(2_i^t) x_j^t + \alpha \Sigma_j p_{ij} \qquad x_j^t = m, p_{ij} = 1$$
$$\Longrightarrow \alpha = \Sigma_t \gamma(2_i^t) m$$
$$\Longrightarrow p_{ij} = \frac{\Sigma \gamma(2_i^t) x_j^t}{-\alpha}$$
$$= \frac{\Sigma_t \gamma(2_i^t) x_j^t}{m \Sigma_t \gamma(2_i^t)}$$

Back to $\frac{\partial L}{\partial p_{ij}} = 0$:

$\Longrightarrow 0 = \Sigma_t \frac{\gamma(2_i^t)}{\pi_i} + \beta$

$= \Sigma_t \gamma(2_i^t) + \pi_i \beta$

$= \Sigma_t \Sigma_i \gamma(2_i) + \Sigma_i \pi_i \beta, \quad \gamma(2_i) = 1, \pi_i = 1$

$\Longrightarrow -\beta = N$

Using $\Sigma_t \gamma(2_i^t) + \pi_i \beta = 0$ from above,

$\pi_i = \frac{\Sigma_t \gamma(2_i^t)}{-\beta}$

$= \frac{N_i}{-\beta}$

$= \frac{N_i}{N}$

$= \frac{\Sigma_{i=1}^N \gamma(z_i^t)}{N}$

To calculate the complete log-likelihood, the hidden variable must be assumed:

$$P(x|2,p) = \prod_{i=1}^{k} P(x|p_i)^{2_i}$$

$$P(2|\pi) = \prod_{i=1}^{k} \pi_i^{2_i}$$

$$\Longrightarrow \log P(x,2|p,\pi) = \sum_{t=1}^{N} \log P(x^t, 2^t|p,\pi)$$

$$= \sum_{t=1}^{N} \log P(x^t|2^t,p) P(2^t|\pi)$$

$$= \sum_{t=1}^{N} \log \prod_{2=1}^{k} (P(x^t|p_i)^{2_i^t} \pi_i^{2_i^t})$$

$$= \sum_{t=1}^{N} \sum_{i=1}^{k} (2_i^t \log \pi_i + 2_i^t \log(P(x^t|p_i))$$

$P(x^t|p_i) = \frac{m!}{x_1^t! \cdots x_n^t!} p_{i1}^{x_1 t} \cdots p_{in}^{x_n^t}$

$\log P(x^t|p_i) = \log\left(\frac{m!}{x^t! \cdots x_n^t!}\right) + \sum_{j=1}^{n} x^t \log p_{ij}$

$log\ (P(x,2|p,\pi) = \sum_{t=1}^{N} \sum_{i=1}^{k} 2_i^t (\log(\pi_i)) + \log(P(x^t|p_i))$

$= \Sigma_{t=1}^N \Sigma_{i=1}^k 2_i^t \left(\log \pi_i + \log \frac{m!}{x_i^t! \cdots x_n^t!} + \Sigma x_j^t \log p_{ij}\right)$

E-Step:

$\Sigma_{t=1}^{N}\Sigma_{\mathrm{i}=1}^{k}(2_i^t|x,p^L)(\log\pi_i+\log(P(x^t|p_i^L)))$

$E(2_i^t|x,p^l)=E(2_1^t|x^t,p^l)$

$=P(2_i=1|x^t,p^l)$

$=\frac{P\left(x^t|p_i^l\right)\pi_i}{\Sigma_{\mathrm{i}}P\left(x^t|p_i^l\right)\pi_i}$

$=\frac{\pi_i\frac{m!}{x_1^t!\cdots x_n^t!}p_{i1}^{x_i^t}\cdots p_{in}^{x_i^t}}{\Sigma_{\gamma=1}^{k}\pi_i\frac{m!}{x_i^t!\cdots x_n^t!}p_{21}^{x_i^t}\cdots p_{in}^{x_i^t}}$

$=\gamma(2_i^t)$

M-Step:

$p^{l+1}=\mathrm{argmax}\Sigma_t\Sigma_i\gamma(2_i^t)\left[\log\pi_i+\log\left(P(x^t|p_i^L)\right)\right]$

$\pi_i=\frac{\Sigma_t r(2_i^t)}{N}$

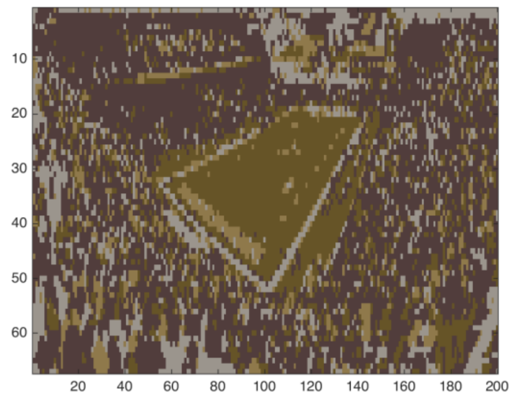
$=\frac{N_i}{N}$

$=\frac{\Sigma_{i=1}^{N}\gamma(z_i^t)}{N}$

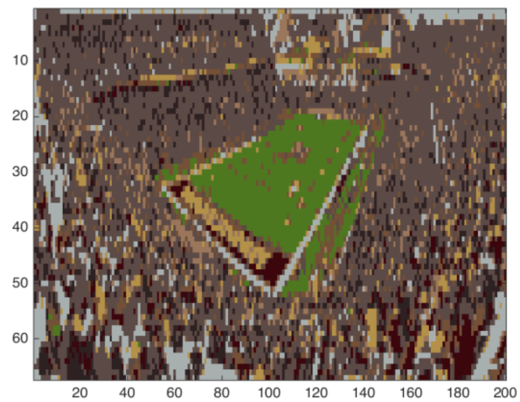
Brian Cooper (5310799)
coope824@umn.edu

Question 2a:

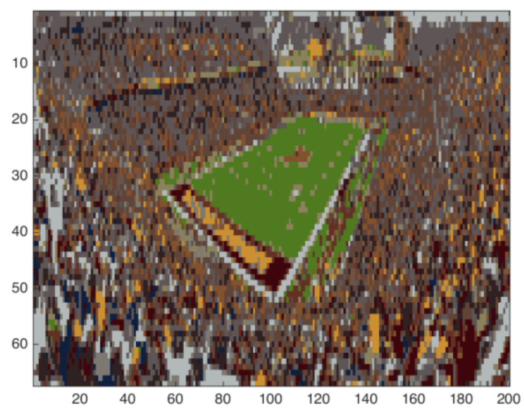
k = 4



k = 8



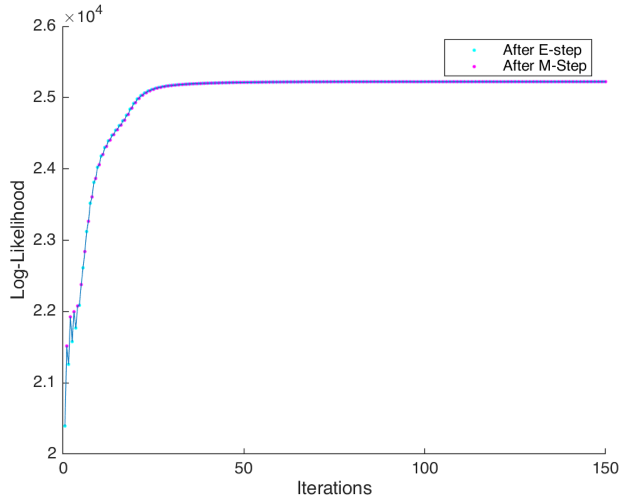
k = 12



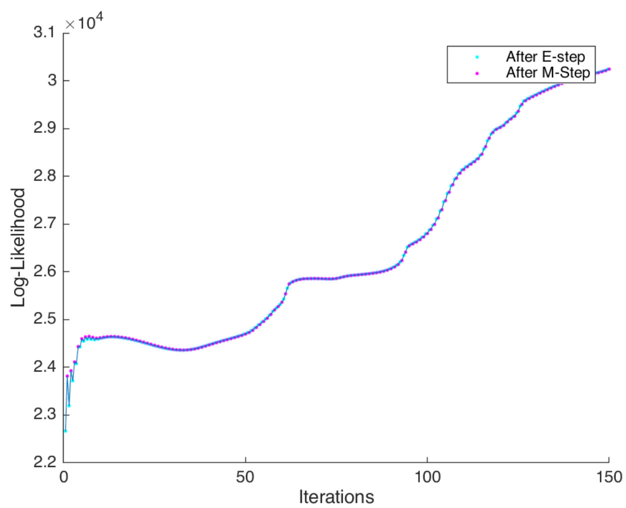
Brian Cooper (5310799)
coope824@umn.edu

Question 2b:

k = 4

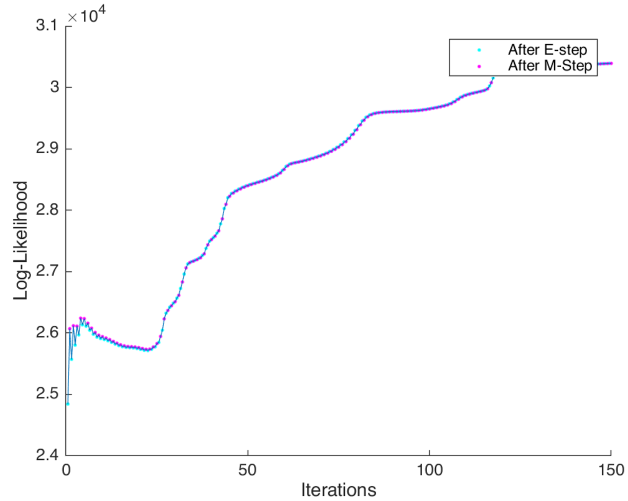


k = 8



Brian Cooper (5310799)
coope824@umn.edu

k = 12



Increasing the value of k also increases the complete log-likelihood value. The complete log-likelihood value and k are positively correlated.

Question 2c:

Attempting to run the EM implementation on "goldy.bmp" failed with the following MATLAB error:

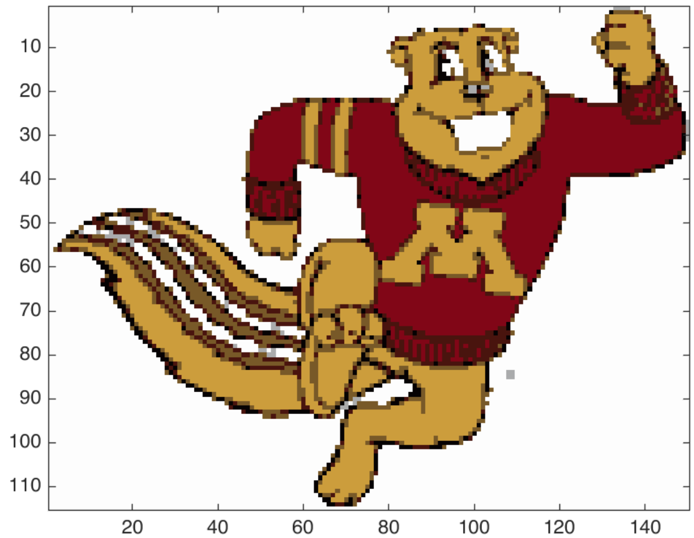
SIGMA must be a square, symmetric, positive definite matrix.

This is because the EM algorithm involves a covariance (sigma) matrix, while k-means does not require the calculation of the covariance matrix and uses Euclidean distance. Specifically, EM requires a positive definite matrix because if it is not, calculating the square root is not possible. These circumstances cause the EM algorithm and the k-means algorithm to behave differently on the image, as can be seen in the images below.

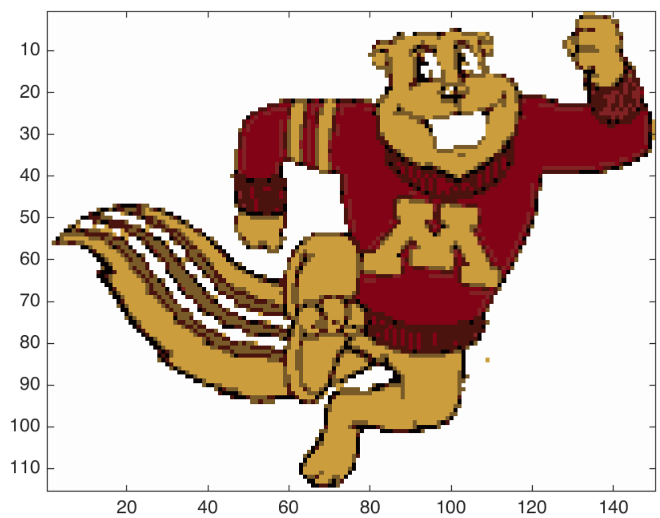
Further, if the covariance matrix is singular, the inverse calculation of such a matrix would be impossible (by definition of a singular matrix). This would cause the EM algorithm to fail.

Brian Cooper (5310799)
coope824@umn.edu

Running the algorithm with the **EM algorithm** produces the following result:



Concurrently, running the algorithm with the built-in **k-means** function produces the following output:



Brian Cooper (5310799)
coope824@umn.edu

Question 2d:

$$\frac{\partial(\frac{-\lambda}{2}\Sigma_{i=1}^{k}\Sigma_{j=1}^{d}(\Sigma_i^{-1})_{jj})}{\partial\Sigma_i^{-1}} = -\frac{\lambda I}{2} \quad \text{(given)}$$

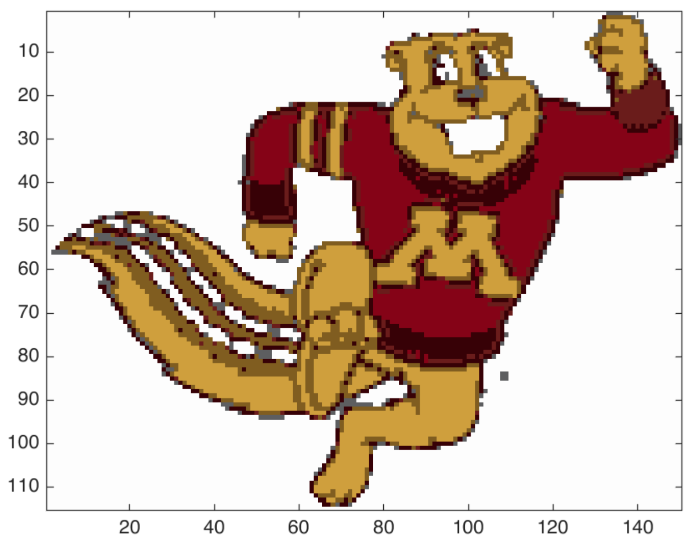
$$f = \Sigma_t\Sigma_i h_i^t(-\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}|x^t - \mu_i|^T\Sigma_i^{-1}(x^t - \mu_i) - \frac{\lambda}{2}\Sigma_j(\Sigma^{-1})_{jj})$$

Set $\frac{\partial f}{\partial \Sigma_i^{-1}} = 0$

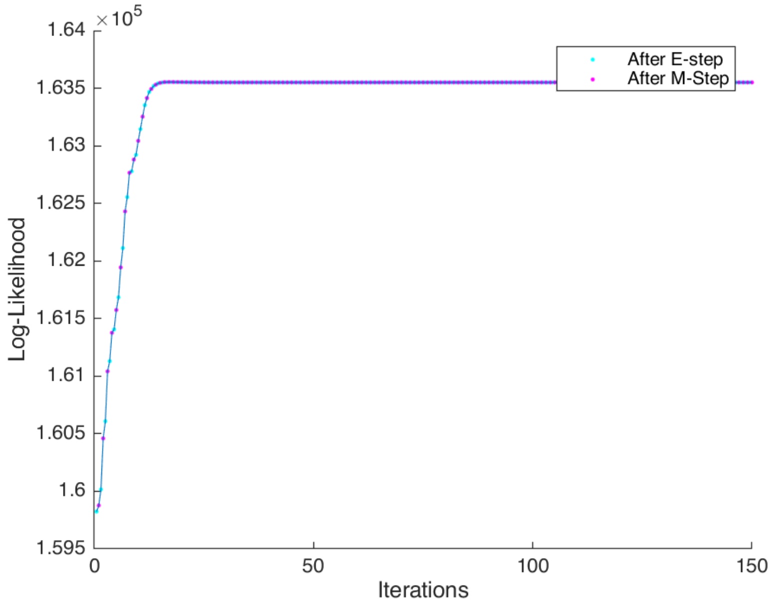
$$\Rightarrow \Sigma_t \frac{h_i^t}{2}\Sigma_i - \frac{1}{2}\Sigma_t h_i^t(x^t - \mu_i)^T(x^t - \mu_i) - \Sigma_t h_i^t\frac{\lambda I}{2} = 0$$

$$\Rightarrow \boxed{\Sigma_i = \frac{\Sigma_t h_i^t((x^t-\mu_i)^T(x^t-\mu_i)+\lambda I)}{\Sigma h_i^t}}$$

Question 2e:



Brian Cooper (5310799)
coope824@umn.edu



Since the modified EM algorithm can no longer be singular, and the covariance matrix is more likely to be positive definite, it successfully operated on the image.