Q1. The probability mass function for a multinomial Distribution for class $c_i$ is given as

$$P(x=(x_1 \dots x_n)|C_i) = P(x=(x_1, \dots x_n)|P_{i1}, \dots P_{in})$$

$$= \frac{m!}{x_1! \dots x_n!} * (P_{i1}^{x_1} \dots P_{in}^{x_n})$$

where $\sum_{j=1}^{n} P_{ij} = 1$ and $\sum_{j=1}^{n} x_j = m$ for all K categories.

The mixture density of K multinomial distributions

$$P(x) = \sum_{i=1}^{K} P(x|C_i)\,P(C_i) = \sum_{i=1}^{K} (\pi_i) \left(\frac{m!}{x_1! \dots x_n!}\right)(P_{i1}^{x_1} \dots P_{in}^{x_n})$$

{where $z_i^t$ is the binary indicator of sample t in cluster i}

$$E(z_i^t) = \frac{\sum_{z_i^t} z_i^t \left( z_i^t \frac{m!}{x_1! \dots x_n!} P_{i1}^{x_1} \dots P_{in}^{x_n}\right)}{\sum_{k=1}^{K} z_k^t \left( z_k \frac{m!}{x_1! \dots x_n!} P_{k1}^{x_1} \dots P_{kn}^{x_2}\right)}$$

$$= \frac{\pi_i \left(\frac{m!}{x_1! \dots x_n!}\right)(P_{i1}^{x_1} \dots P_{in}^{x_n})}{\sum_{k=1}^{K} \pi_k \left(\frac{m!}{x_1! \dots x_n!}\right)(P_{k1}^{x_1} \dots P_{kn}^{x_n})}$$

$$E(z_i^t) = \frac{\pi_i (P_{i1}^{x_1} \dots P_{in}^{x_n})}{\sum_{k=1}^{K} \pi_k (P_{k1}^{x_1} \dots P_{kn}^{x_n})} = \cancel{\gamma}\ \gamma(z_i^t)$$

we now have a ~~functio~~ formula to decide the responsibility $(\gamma(z_i^t))$

~~we know will derive the log likelihood estimation is~~

we can now derive the complete log likelihood function and formulas to increment our parameters.

$$E_c[\ln P(x, c \mid p_1, p_2 \ldots p_n, \pi)]$$

$$\mathcal{L} = \sum_{t=1}^{N} \sum_{i=1}^{K} \gamma(z_i^t) \left( \ln(\pi_i) + \ln\left( \frac{m!}{x_1^t! \ldots x_n^t!} p_{i1}^{x_1^t} \ldots p_{in}^{x_n^t} \right) \right)$$

Partially deriving the equation w.r.t $p_{ij}$

$$\frac{\partial \mathcal{L}}{\partial p_{ij}} = \frac{\partial}{\partial p_{ij}} \left( \left\{ \sum_{t=1}^{N} \sum_{i=1}^{K} \gamma(z_i^t) \left[ \ln \pi_i + \ln\left( \frac{m!}{x_1^t! \ldots x_n^t!} p_{i1}^{x_1^t} \ldots p_{in}^{x_n^t} \right) \right] \right. \right.$$

$$\left. + \lambda \left( \sum p_{ij} - 1 \right) \right)$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial p_{ij}} = \sum_{t=1}^{N} \gamma(z_i^t) \frac{x_j^t}{p_{ij}} + \lambda$$

$$0 = \sum_{t=1}^{N} \gamma(z_i)^t \frac{x_j}{p_{ij}} + \lambda.$$

$$\lambda = \sum_{t=1}^{N} \gamma(z_i^t) \left( \frac{(x_j)^t}{p_{ij}} \right)$$

$$p_{ij} = -\left( \frac{\sum_{t=1}^{N} y(z_i^t) + (x_j)^t}{\lambda} \right)$$

We have that $\sum_{j=1}^{n} p_{ij} = 1$

$$\sum_{j=1}^{n} p_{ij} = 1 = -\sum_{j=1}^{n} \left( \frac{\sum_{t=1}^{N} y(z_i^t)\, x_j^t}{\lambda} \right)$$

$$\lambda = -m N_i$$

$$p_{ij} = \frac{\sum_{t=1}^{N} y(z_i^t)\,(x_j)^t}{m N_i}$$

we now ~~deri~~ get the partial derivative of $l$ wrt to $\pi_i$

$$\frac{\partial l}{\partial \pi_i} = \frac{\partial}{\partial \pi_i} \left( \sum_{t=1}^{N} \sum_{i=1}^{k} y(z_i^t) \left\{ \ln \pi_i + \ln \left( \frac{M!}{x_1^t! \cdots x_n^t!} \right) p_{i1}^{x_1^t} \cdots p_{in}^{x_n^t} \right\} \right.$$

$$\left. + \lambda \left( \sum_{i=1}^{k} \pi_i - 1 \right) \right\} \Big)$$

Taking the derivative $= 0$.

$$\frac{\partial l}{\partial \pi_i} = \sum_{t=1}^{N} \left( \frac{y(z_i^t)}{\pi_i} \right) + \lambda = 0$$

$$\Rightarrow \quad -\lambda = \frac{\sum_{t=1}^{N} ( y(z_i^t) )}{\pi_i}$$

$$\Rightarrow \quad \pi_i = -\frac{\sum_{t=1}^{N} \left( y(z_i^t) \right)}{\lambda} = -\frac{N_i}{\lambda}$$

we know that $\sum_{i=1}^{k} \pi_i = 1$.

by substitution the value of $\pi_i$ we get.

$$1 = -\sum_{i=1}^{k} \left( \frac{\sum_{t=1}^{N} (y(z_i^t))}{\lambda} \right)$$

$$1 = -\left(\frac{1}{\lambda}\right) N$$

$$\cancel{N = -\lambda} \quad \lambda = -N$$

using this in the equation for $\pi_i$ we get

$$\pi_i = \frac{-N_i}{\lambda} = \frac{-N_i}{-N} = \left(\frac{N_i}{N}\right)$$

$$\pi_i = \frac{N_i}{N}$$

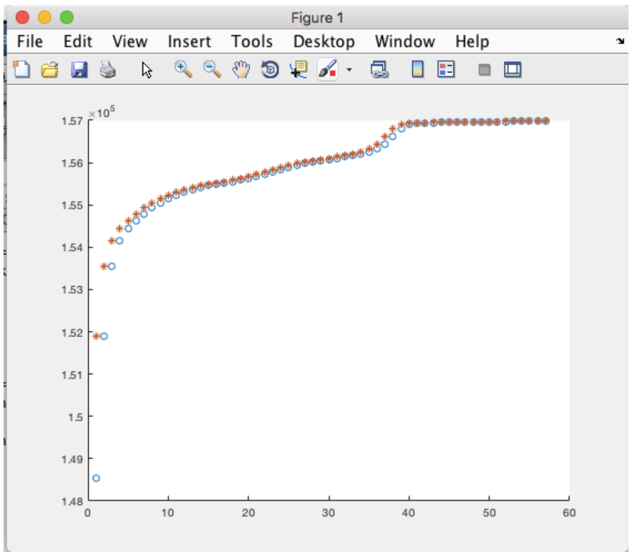
$$P_{ij} = \frac{\sum_{t=1}^{N} y(z_i^t)\,(x_j^t)}{m N_i}$$

and we have our expectation function

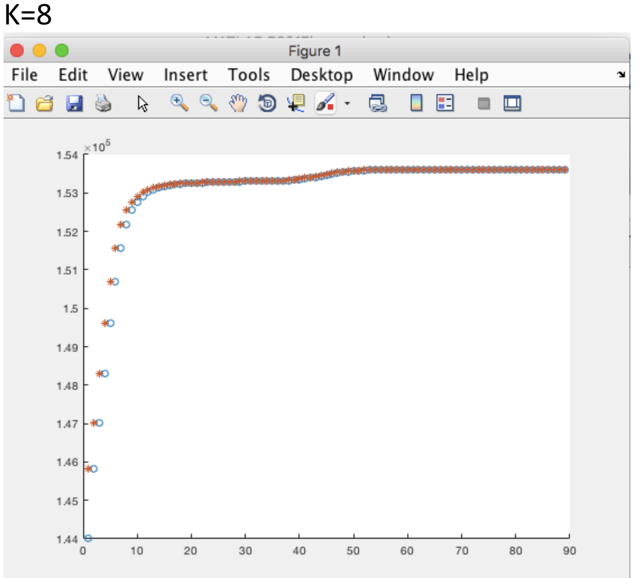
$$y(z_i^t) = \frac{\pi_i \left(P_{i1}^{x_1} \cdots P_{in}^{x_n}\right)}{\sum_{k=1}^{k} \pi_k \left(P_{k1}^{x_1} \cdots P_{kn}^{x_n}\right)}$$

Q2(b) On running our initial EM implementation on the 'stadium.bmp' file for k=4,8 and 12 we get the following log likelihood function plots.
Here the red points are log likelihood values calculated after the maximization step. The blue points are the log likelihood values calculated after the expectation step.

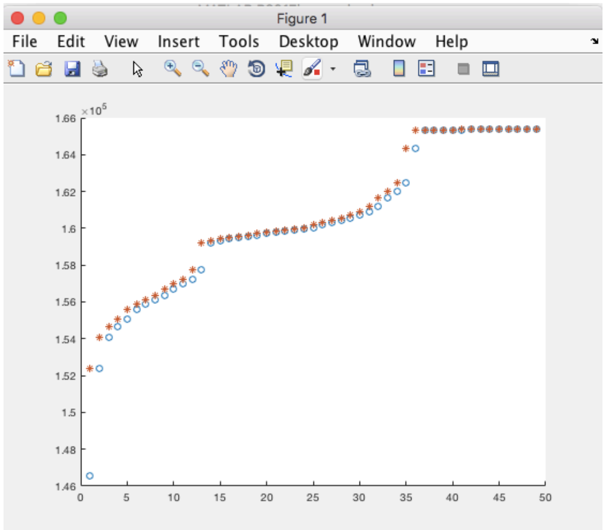
K=4




K=8

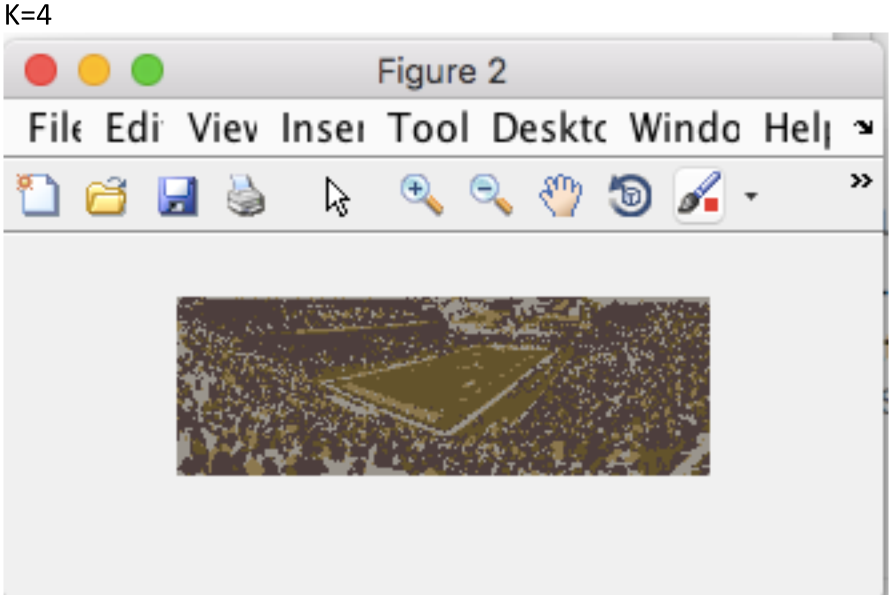


K=12

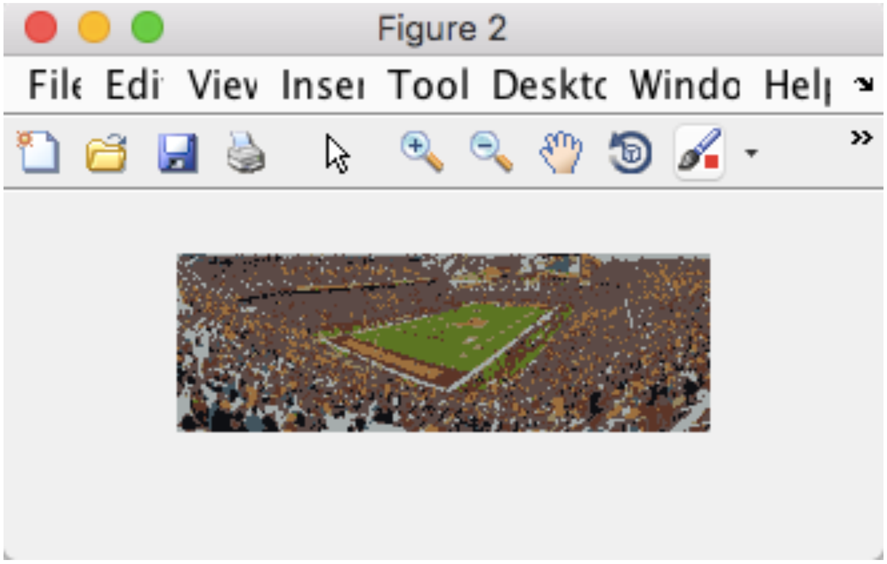



All three plots follow the general characteristic of increasing with each step and then plateauing at the maximum value where the two log likelihood values of the iteration have a different lesser than 0.1 and can be considered as almost equal. The actual curvature of the plots are also dependent on what initial random clusters are assumed. If they are close to the clusters that the EM algorithm estimates the log values would converge more quickly.
Also included are the images generated by using the mean values of the cluster as the colors of the group pixels.

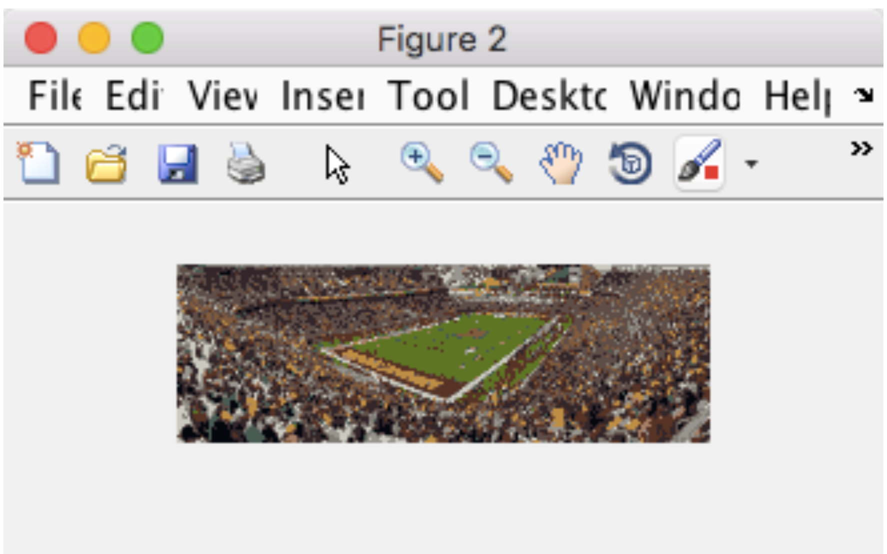
K=4



K=8



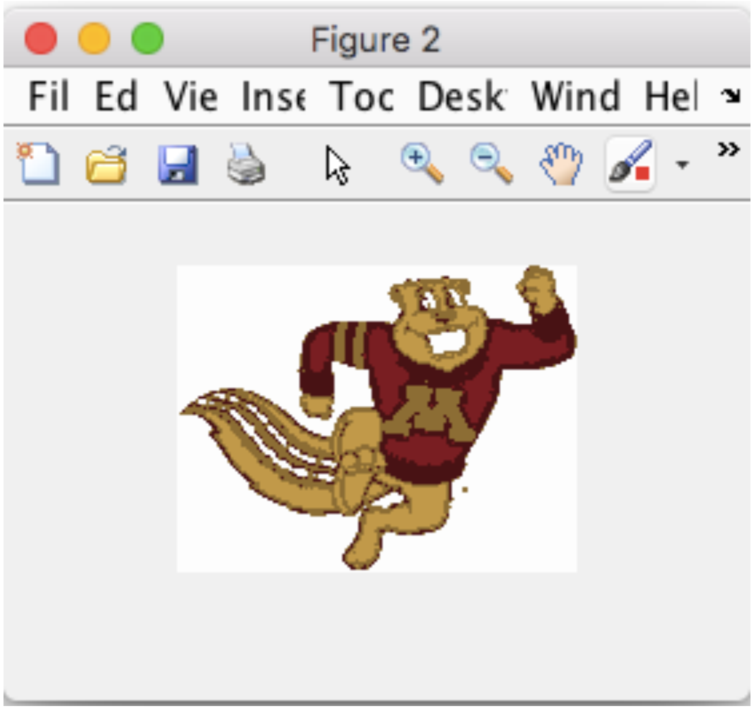

K=12



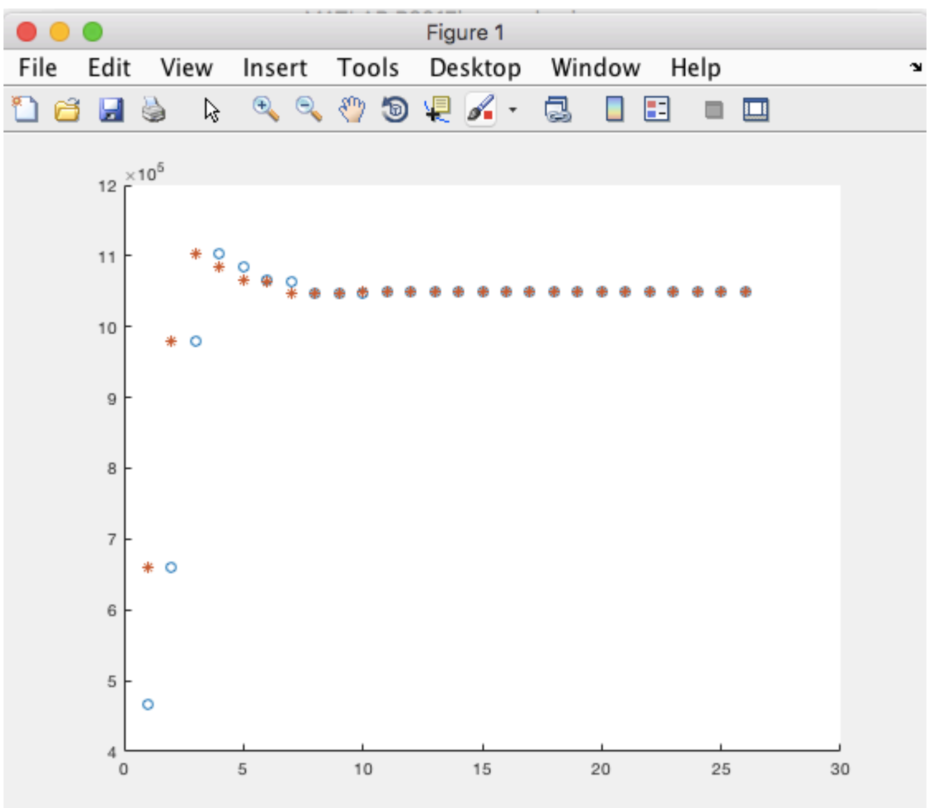

Q1c

We run our  implementation of EM on ‘goldy.bmp’ and get the following plots and recreated image.

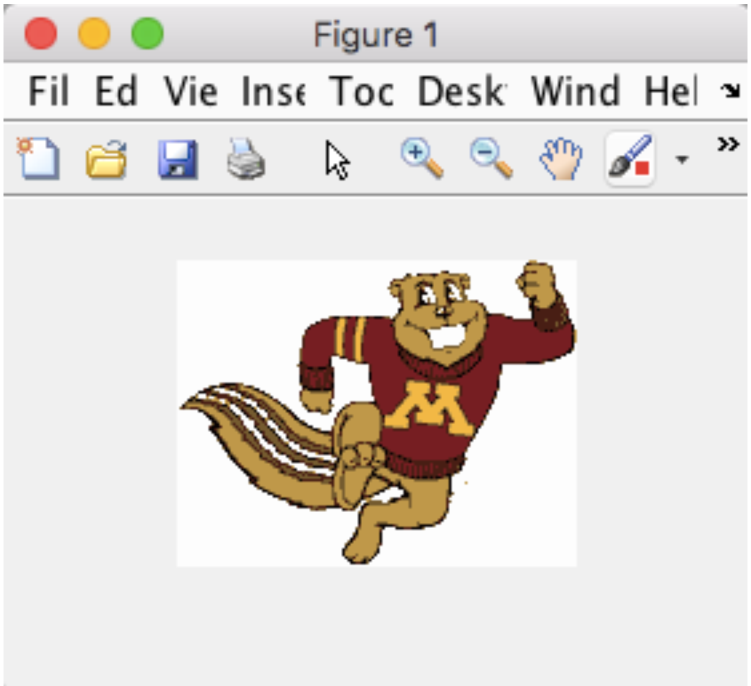
Goldy recreated.




Log likelihood plot



When we run the k-means algorithm on the same image we get the following




On comparing the recreation of the image by EM and k-means with the same numbers of clusters we notice that the boundaries between the different clusters(in this case colors) are more pronounced in the k-means approach due to the hard clustering. Due to this the “M” in the center is distinct from the red shirt around it. However the EM approach is affected by the red pigmentation around the “M” and therefore the color itself has more “red’ making it almost brownish.

Q2D

Q2(d) The expectation of the log likelihood function with regularization looks like the following.

$$\mathcal{L} = \sum_{t=1}^{N} \sum_{i=1}^{K} \gamma(z_i^t) \left( \log \pi_i + \log \frac{1}{(2\pi)^{d/2} (\Sigma_i)^{\frac{1}{2}}} e^{1/2 (x^t - \mu_i) \Sigma^{-1} (x^t - \mu_i)'} \right)$$

$$- \frac{\lambda}{2} \sum_{i=1}^{K} \sum_{j=1}^{d} (\Sigma_i^{-1})_{jj}$$

On taking the partial derivative of $\mathcal{L}$ w.r.t to $\Sigma^{-1}$ and equating that to 0 we have

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \sum_{t=1}^{N} (z_i^t) \left( \frac{\Sigma_i}{2} - \frac{1}{2} (x^t - \mu_i) (x^t - \mu_i)^T \right) - \frac{\lambda I}{2}$$

$$\frac{\partial L}{\partial \Sigma^{-1}} = 0$$

$$\Rightarrow \sum_{t=1}^{N} (z_i^t) \left( \frac{\Sigma_i}{2} - \frac{1}{2} (x^t - \mu_i) (x^t - \mu_i)^T \right) - \frac{\lambda I}{2}$$

$$\frac{\Sigma N_i}{2} = \frac{1}{2} \sum_{i=1}^{N} \gamma(z_i^t) \left( \frac{\Sigma_i}{2} - \frac{1}{2} (x^t - \mu_i) (x^t - \mu_i)^T \right) + \frac{\lambda I}{2}$$

$$\cancel{\Sigma = \frac{1}{N} \sum_{i=1}^{N} \gamma(z_i^t) (x_i^t - \mu_i) (x_i^t - \mu}$$

$$\Sigma = \frac{1}{N} \sum_{t=1}^{N} \gamma(z_i^t) (x^t - \mu_i) (x^t - \mu_i)^T + \frac{\lambda I}{N_i}$$

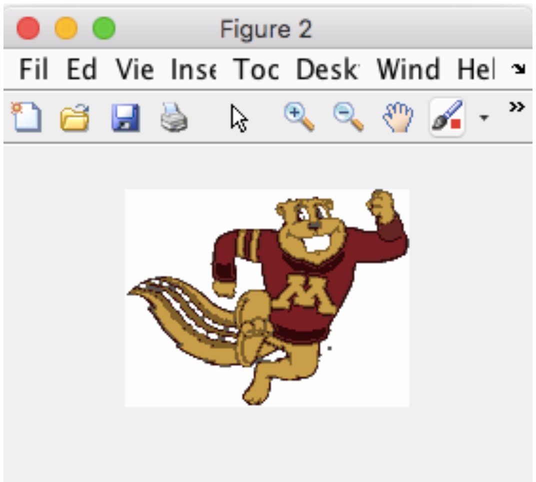
Therefore we have

$$\Sigma = \frac{1}{N} \sum_{i=1}^{N} \gamma(z_i^t) (x^t - \mu_i) (x^t - \mu_i)^T + \frac{\lambda I}{N_i}$$

Scanned with CamScanner

Q2E

By including the regularization term into the model allows for the values of log likelihood values to converge more quickly are the changes to the covariance matrix of each cluster mean is reduced. This leads to more general classification which makes more a stable model and reduces the risk of overtraining. The results of running the EM model on “goldy.bmp” with k=7 and regularization turned on is given below.




Log likelihood plot.