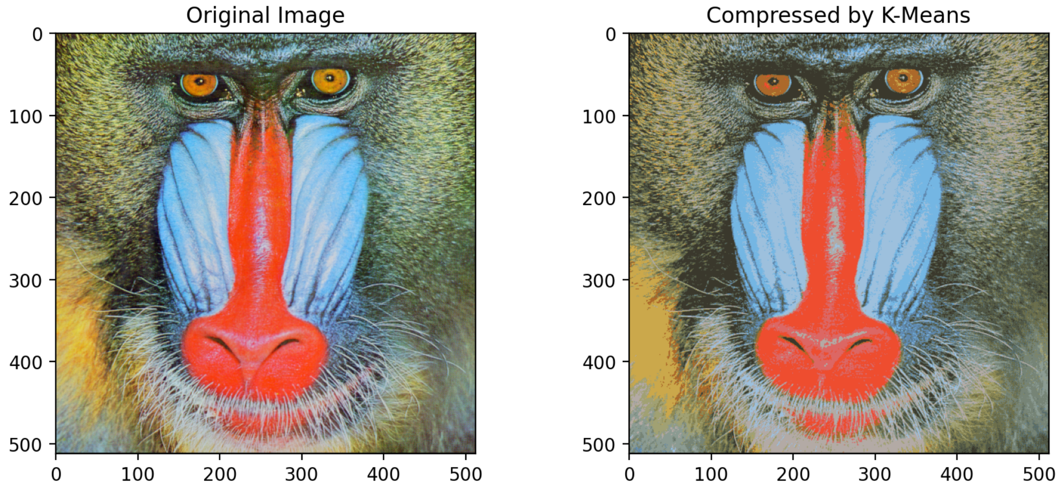
1.1 (a) AG

(b)



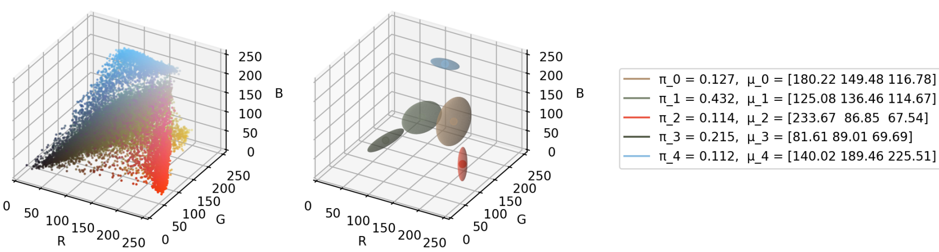
Mean error is 11.9358298.

(c) Original image needs 24 bits and there are 16 clusters for compressed image needs $\log_2(16) = 4$ bits. Thus, the factor will be $\frac{24}{4} = 6$

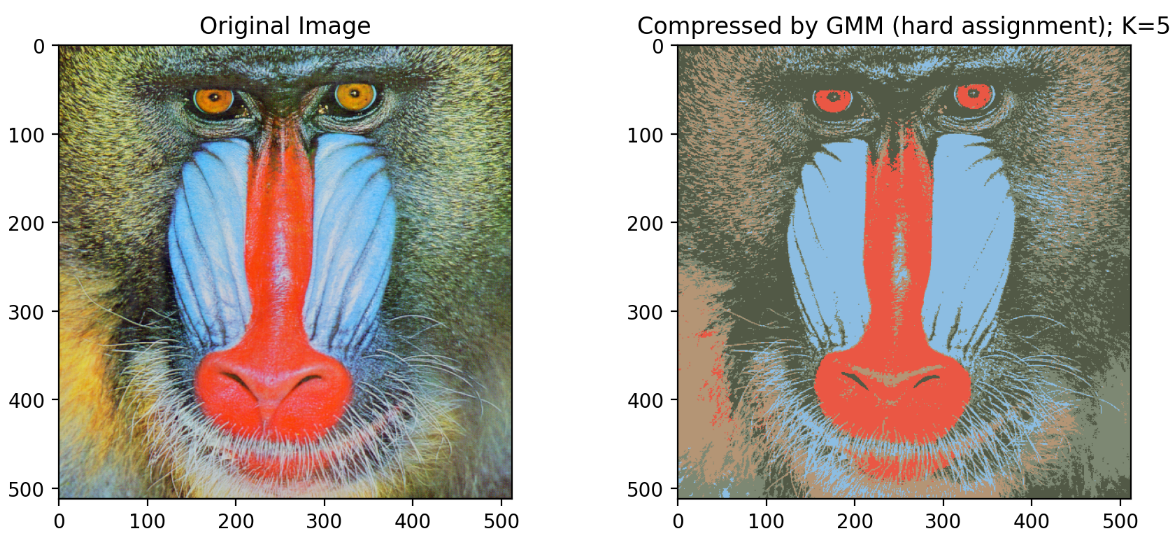
1.2 (d)

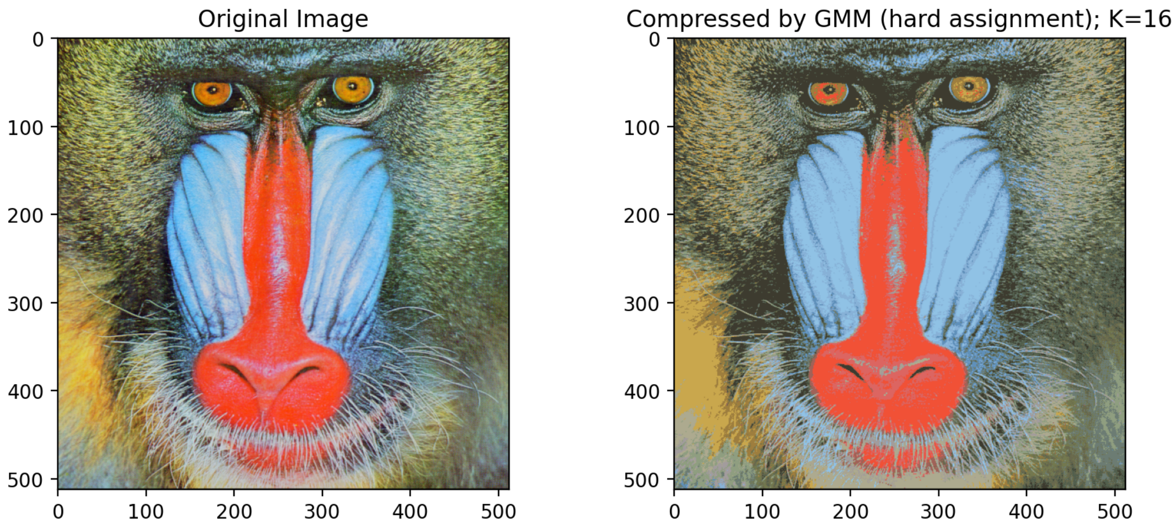
(e)

```
pi = array([0.13, 0.43, 0.11, 0.21, 0.11])
mu = array([[180.22, 149.48, 116.78],
       [125.08, 136.46, 114.67],
       [233.67,  86.85,  67.54],
       [ 81.61,  89.01,  69.69],
       [140.02, 189.46, 225.51]])
sigma = array([[[ 521.36,  290.61, -210.66],
        [ 290.61, 1047.03,  842.88],
        [-210.66,  842.88, 2511.74]],

       [[ 742.09,  526.24,   40.68],
        [ 526.24,  607.66,  454.3 ],
        [  40.68,  454.3 , 1226.68]],

       [[ 129.72, -108.61, -238.98],
        [-108.61,  259.27,  460.1 ],
        [-238.98,  460.1 , 1035.17]],

       [[ 459.17,  518.45,  324.74],
        [ 518.45,  692.35,  486.85],
        [ 324.74,  486.85,  463.54]],

       [[ 612.96,  146.43,  -90.39],
        [ 146.43,   89.24,   32.62],
        [ -90.39,   32.62,   82.26]]])
```



(f)





2.(a) $J = \sum_{i=1}^{l} \log p(x^{(i)}, y^{(i)}) + \lambda \sum_{l+1}^{l+u} \log p(x^{(i)})$

$$\Rightarrow E_{q_i}[\log p(x^{(i)}, y^{(i)})] = \sum_{j=0}^{1} q_i(y^{(i)}=j) \log \frac{p(x^{(i)}, y^{(i)}=j)}{q_i(y^{(i)}=j)} + \log q_i(y^{(i)}=j)$$

$\Rightarrow$ apply Jensen's inequality.

$$\log \sum_{j=0}^{1} q_i(y^{(i)}=j) \frac{p(x^{(i)}, y^{(i)}=j)}{q_i(y^{(i)}=j)} \geq \sum_{j=0}^{1} q_i(y^{(i)}=j) \frac{p(x^{(i)}, y^{(i)}=j)}{q_i(y^{(i)}=j)}$$

$$\Rightarrow \log p(x^{(i)}) \geq \sum_{j=0}^{1} q_i(y^{(i)}=j) \log \frac{p(x^{(i)}, y^{(i)}=j)}{q_i(y^{(i)}=j)}$$

Summing ~~this~~ inequality over all unlabeled examples and incorporating the labeled data likehood, we have:

$$L(\mu, \Sigma, \phi) = \sum_{i=1}^{l} \log p(x^{(i)}, y^{(i)}) + \lambda \sum_{i=l+1}^{l+u} \sum_{j=0}^{1} Q_{ij} \log \frac{p(x^{(i)}, y^{(i)}=j)}{q_i(y^{(i)}=j)}$$

(where $Q_{ij} = q_i(y^{(i)}=j)$).

(b) $$P(y^{(i)}=j \mid x^{(i)}) = \frac{P(x^{(i)} \mid y^{(i)}=j)\,P(y^{(i)}=j)}{P(x^{(i)})}$$

$$\Rightarrow P(y^{(i)}=j \mid x^{(i)}) = \frac{P(x^{(i)} \mid \mu_j, \Sigma_j)\,P(y^{(i)}=j \mid \phi)}{\sum_{k=0}^{1} P(x^{(i)} \mid \mu_k, \Sigma_k)\,P(y^{(i)}=k \mid \phi)}$$

$$\Rightarrow Q_{ij} = \frac{\phi^{j}(1-\phi)^{1-j}\,N(x^{(i)}; \mu_j, \Sigma_j)}{\phi N(x^{(i)}; \mu_j, \Sigma_j) + (1-\phi) N(x^{(i)}; \mu_0, \Sigma_0)}$$

(c) The likelihood for μ_k is:

$$\sum_{i=1}^{l+u} Q_{ik} \log N(x^{(i)}; \mu_k, \Sigma_k) \xrightarrow[\text{Gaussian}]{\text{multivariance}}$$

$$\rightarrow \sum_{i=1}^{l+u} Q_{ik}\left[-\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x^{(i)}-\mu_k)^T \Sigma_k^{-1}(x^{(i)}-\mu_k)\right].$$

$$= \sum_{i=1}^{l+u} Q_{ik}(x^{(i)}-\mu_k)^T \Sigma_k^{-1}(x^{(i)}-\mu_k)$$

$$\Rightarrow \frac{\partial}{\partial \mu_k} \sum_{i=1}^{l+u} Q_{ik}(x^{(i)}-\mu_k)^T \Sigma_k^{-1}(x^{(i)}-\mu_k) = 0$$

$$\Rightarrow \sum_{i=1}^{l+u} Q_{ik} \Sigma_k^{-1}(x^{(i)}-\mu_k) = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{i=1}^{l+u} Q_{ik} x^{(i)}}{\sum_{i=1}^{l+u} Q_{ik}}$$

(d) $$L(\phi) = \sum_{i=1}^{l} \log P(y^{(i)}) + \lambda \sum_{i=l+1}^{l+u} \sum_{j=0}^{1} \log P(y^{(i)} = j)$$

$$= \sum_{i=1}^{l} \log P(y^{(i)}) + \lambda \sum_{i=l+1}^{l+u} Q_{i1} \log \phi + Q_{i0} \log(1-\phi)$$

$$= \sum_{i=1}^{l} y^{(i)} \log \phi + (1-y^{(i)}) \log(1-\phi) +$$

$$\lambda \sum_{i=l+1}^{l+u} Q_{i1} \log \phi + Q_{i0} \log(1-\phi)$$

$$\Rightarrow \frac{\partial L}{\partial \phi} = \sum_{i=1}^{l} \frac{y^{(i)}}{\phi} - \frac{1-y^{(i)}}{1-\phi} + \lambda \sum_{i=l+1}^{l+u} \frac{Q_{i0}}{1-\phi} = 0$$

$$\Rightarrow \phi = \frac{\sum_{i=1}^{l} y^{(i)} + \lambda \sum_{i=l+1}^{l+u} Q_{i1}}{l + \lambda u}$$

(e) $$\Sigma_k = \frac{\sum_{i=1}^{l+u} Q_{ik} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T}{\sum_{i=1}^{l+u} Q_{ik}}$$

3(a) $\frac{1}{N}\sum_{i=1}^{N}\|x^{(i)}-UU^Tx^{(i)}\|^2=\frac{1}{N}\|X-UU^TX\|^2=tr\left(\frac{1}{N}(X-UU^TX)^T(X-UU^TX)\right)$

$$\Rightarrow\ = tr\left(\frac{1}{N}X^TX\right)-tr\left(\frac{1}{N}X^TUU^TX\right)$$

$$= tr\left(\frac{1}{N}XX^T\right)-tr\left(U^T\left(\frac{1}{N}XX^T\right)U\right)$$

$$= tr(S)-tr(U^TSU)$$

$$= \sum_{i=1}^{d}\lambda_i-\sum_{i=1}^{K}u_i^TSu_i \quad , \text{(} \lambda_i \text{ are the eigenvalues of } S\text{)}$$

$$\Rightarrow\ \underset{\forall j,\|u_j\|_2=1}{\arg\min}\ \frac{1}{N}\sum_{n=1}^{N}\left\|x^{(n)}-\sum_{i=1}^{K}u_iu_i^Tx^{(n)}\right\|^2$$

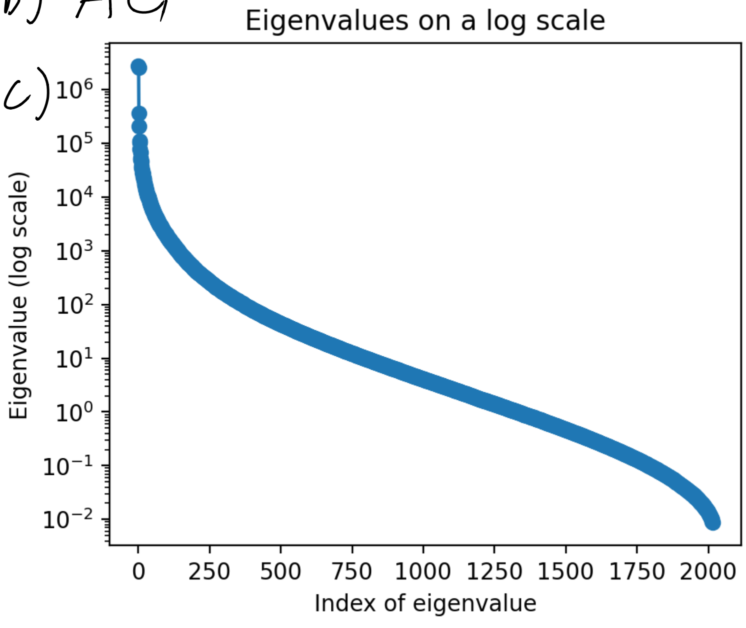
$$= \underset{\forall j,\|u_j\|_2=1}{\arg\min}\ \sum_{i=1}^{d}\sum_{i=1}^{d}\lambda_i-\sum_{i=1}^{K}u_i^TSu_i$$

$$= \underset{\forall j,\|u_j\|_2=1}{\arg\max}\ \sum_{i=1}^{K}u_i^TSu_i$$

$$\Rightarrow\ \min_{\forall j,\|u_j\|_2=1}\ \sum_{i=1}^{d}\lambda_i-\sum_{i=1}^{K}u_i^TSu_i=\sum_{i=1}^{d}\lambda_i-\sum_{i=1}^{K}\lambda_i=\sum_{i=K+1}^{d}\lambda_i$$

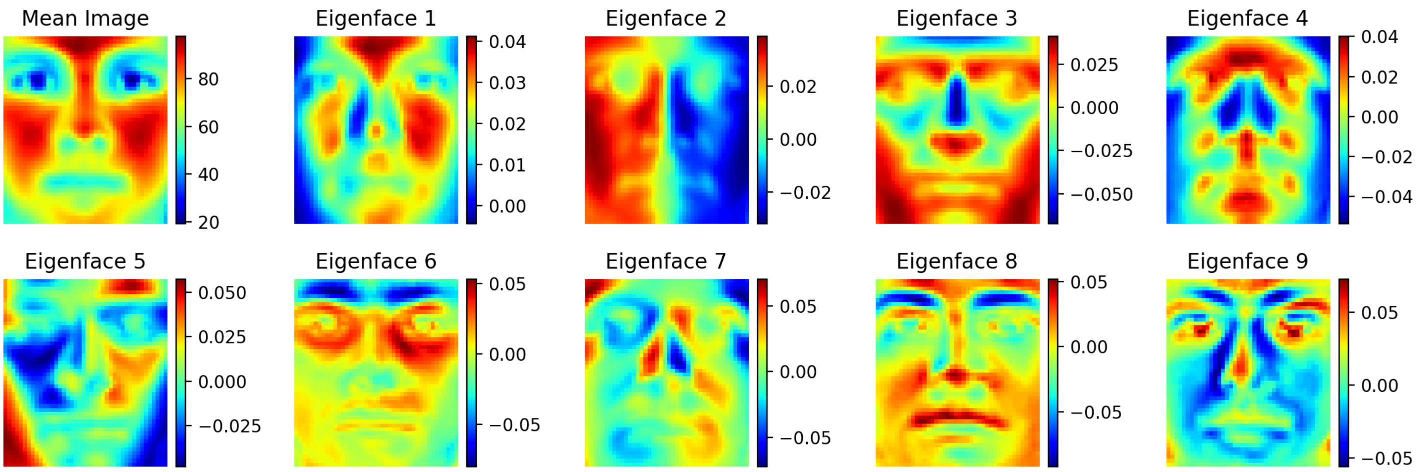
(b) AG

(c)




```
.95964519  365515.29797577  211149.83657024
.79567702   78512.9353901    67032.06261281
.08114141]
```

(d)



Eigenface 3 could show the shape of nose.

Eigenface 1 could show the shape of eyeball

eigenface 8 could show the shape of lip.

And 2 for left and right face.

(e) 95% : # components is 43, reduction is 97.867%

99% : # components is 167, reduction is 91.716%

4. (a) AG

(b)

```
Separating tracks ...
working on alpha = 0.1
working on alpha = 0.1
working on alpha = 0.1
working on alpha = 0.05
working on alpha = 0.05
working on alpha = 0.05
working on alpha = 0.02
working on alpha = 0.02
working on alpha = 0.01
working on alpha = 0.01
working on alpha = 0.005
working on alpha = 0.005
working on alpha = 0.002
working on alpha = 0.002
working on alpha = 0.001
working on alpha = 0.001
W solution:
 [[ 72.15081922  28.62441682  25.91040458 -17.2322227  -21.191357  ]
 [ 13.45886116  31.94398247  -4.03003982 -24.0095722   11.89906179]
 [ 18.89688784  -7.80435173  28.71469558  18.14356811 -21.17474522]
 [ -6.0119837   -4.15743607  -1.01692289  13.87321073  -5.26252289]
 [ -8.74061186  22.55821897   9.61289023  14.73637074  45.28841827]]
CPU times: user 1min, sys: 22.2 s, total: 1min 22s
Wall time: 16.1 s
```

5. (a) $\log P_\theta(x|y) = \log\int P_\theta(x|z,y)\, P(z|y)\, dz$

$\therefore\ \log P_\theta(x|y) = \log\int P_\theta(x|z,y)\,\frac{q_\phi(z|x,y)}{q_\phi(z|x,y)}\, P(z|y)\, dz$

Apply Jensen's inequality

$\Rightarrow$

$\log P_\theta(x|y) \geq \int q_\phi(z|x,y) \log\left(\frac{P_\theta(x|z,y)P(z|y)}{q_\phi(z|x,y)}\right) dz$

$\Rightarrow \log P_\theta(x|y) \geq E_{q_\phi(z|x,y)}\left[\log P_\theta(x|z,y) - E_{q_\phi(z|z,y)}\left[\log \frac{q_\phi(z|x,y)}{P(z|y)}\right]\right]$

$\Rightarrow \log P_\theta(x|y) \geq E_{q_\phi(z|x,y)}\left[\log P_\theta(x|z,y) - D_{KL}(q_\phi(z|x,y)\,||\,p(z|y))\right]$

(b) The KL-divergence between two multivariate Gaussian distributions:

$D_{KL}(N_0||N_1) = \frac{1}{2}\left(tr(\Sigma_1^{-1}\Sigma_0) + (\mu_1-\mu_0)^T\Sigma_1^{-1}(\mu_1-\mu_0) - k + \log\left(\left|\frac{\Sigma_1}{\Sigma_0}\right|\right)\right)$

In this equation, k is dimension, and $P_\theta(z|y) \sim N(0, I) \Rightarrow \begin{cases}\mu_1 = 0\\ \Sigma_1 = I\end{cases}$

$\Rightarrow D_{KL}(q_\phi||P_\theta) = \frac{1}{2}\left(\sum_{j=1}^{M}\sigma_j^2 + \mu_j^2 - \log(\sigma_j^2) - 1\right) \Rightarrow \Sigma_1 = I$

$\Rightarrow D_{KL}(q_\phi||P_\theta) = \frac{1}{2}\left(\sum_{j=1}^{M}\sigma_j^2 + \mu_j^2 - \log(\sigma_j^2 - 1)\right)$

(c)