1.(b) The original picture is 256 RGB, so each pixel requires

$\log_2 256 \times 3 = 8 \times 3 = 24$ bits of memory

For the compressed image, each pixel requires

$\log_2 16 = 4$ bits of memory

The overal compression factor is $24/4 = 6$

2.(a)

$$\ell_{semi\_sup}(\theta^{(t+1)}) = \alpha \ell_{sup}(\theta^{(t+1)}) + \ell_{unsup}(\theta^{(t+1)})$$

$$\geq \alpha \ell_{sup}(\theta^{(t+1)}) + \sum_{i=1}^{n} \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})}$$

$$\geq \alpha \ell_{sup}(\theta^{(t)}) + \sum_{i=1}^{n} \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{Q_i^{(t)}(z^{(i)})}$$

$$= \alpha \ell_{sup}(\theta^{(t)}) + \ell_{unsup}(\theta^{(t)})$$

Where the first inequality holds because of Jensen's inequality

the second inequality holds because $\theta^{(t+1)}$ is obtained by

$$\arg\max_{\theta} \left[\sum_{i=1}^{n} \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i^{(t)}(z^{(i)})} + \alpha \ell_{sup}(\theta) \right] \text{ at } t+1.$$

and the equality holds because we choose.

$$Q_i^{(t)}(z^{(i)}) = p(z^{(i)} | x^{(i)}; \theta^{(t)}) \text{ for the E step at } t$$

Thus $\ell_{semi\text{-}sup}(\theta^{(t+1)}) \geq \alpha \ell_{sup}(\theta^{(t)}) + \ell_{unsup}(\theta^{(t)}) = \ell_{semi\text{-}sup}(\theta^{(t)})$

2.(b)

At E step, for each i, set

$$w_j^{(i)} = Q_i(z^{(i)}=j) = P(z^{(i)}=j \mid x^{(i)}; \phi, \mu, \Sigma)$$

$$= \frac{P(x^{(i)} \mid z^{(i)}=j; \mu_j, \Sigma_j)\,\phi_j}{\sum_{j=1}^{k} P(x^{(i)} \mid z^{(i)}=j; \mu_j, \Sigma_j)\,\phi_j}$$

Where $P(x^{(i)} \mid z^{(i)}=j; \mu_j, \Sigma_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_j)^T \Sigma_j^{-1} (x^{(i)}-\mu_j)\right)$

2.

2.(c)

Since M step becomes

$$\arg\max_{\phi,\mu,S} \sum_{i=1}^{n}\sum_{j=1}^{k} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \phi, \mu, S)}{Q_i^{(t)}(z^{(i)})} + \sum_{i=1}^{\tilde{n}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \phi, \mu, S)$$

$$= \arg\max_{\phi,\mu,S} \left(\sum_{i=1}^{n}\sum_{j=1}^{k} w_j^{(i)} \log \frac{\frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_j)^T \Sigma_j^{-1}(x^{(i)}-\mu_j)\right)\phi_j}{w_j^{(i)}} \right.$$

$$\left. + \alpha \sum_{i=1}^{\tilde{n}}\sum_{j=1}^{k} \log \frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(\tilde{x}^{(i)}-\mu_j)^T \Sigma_j^{-1}(\tilde{x}^{(i)}-\mu_j)\right) 1\{z^{(i)}=j\}\phi_j \right)$$

by taking derivative w.r.t μ_j and then setting the derivative to zero

$$\Rightarrow \sum_{i=1}^{n} w_j^{(i)}\left(\Sigma_j^{-1}x^{(i)} - \Sigma_j^{-1}\mu_j\right) + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}\left(\Sigma_j^{-1}\tilde{x}^{(i)} - \Sigma_j^{-1}\mu_j\right) = 0$$

$$\Rightarrow \mu_j = \frac{\sum_{i=1}^{n} w_j^{(i)} x^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)} \tilde{x}^{(i)}}{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}$$

by taking the derivative w.r.t. S_j and then setting the derivative to zero

$$\Rightarrow \Sigma_j = \frac{\sum_{i=1}^{n} w_j^{(i)}(x^{(i)}-\mu_j)(x^{(i)}-\mu_j)^T + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}(\tilde{x}^{(i)}-\mu_j)(\tilde{x}^{(i)}-\mu_j)^T}{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}$$

also from the handout of PS-5, we have already seen how to derive ϕ_j as

$$\phi_j = \frac{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}{n + \alpha\tilde{n}}$$

2(f). Semi-supervised EM takes less iterations to converge, is more stable, and has higher quality of assignments compared to unsupervised EM.