Q.1 M-Step.

1. The objective function we want to optimize,

$$Q(\theta,\theta^{old}) = \sum_{i=1}^{N}\left\{\sum_{z^i=1}^{K} P(z^i|x^i,\theta^{old}) * \log\left(P(x^i,z^i|\theta)\right)\right\}.$$

$$= \sum_{i=1}^{N}\left\{\sum_{z^i=1}^{K} P(z^i|x^i,\theta^{old}) \cdot \log\left(P(z^i)\cdot P(x^i|\theta)\right)\right\}.$$

$$= \sum_{i=1}^{N}\left\{\sum_{z^i=1}^{K} P(z^i|x^i,\theta^{old}) \cdot \log\left(\pi_{z_i} P(x^i|\theta)\right)\right\}.$$

$$= \sum_{i=1}^{N}\sum_{z^i=K}^{K} P(z^i_{=k}|x^i,\theta^{old}) \log(\pi_{z_i}) + \sum_{i=1}^{N}\sum_{k=1}^{K} P(z^i_{=k}|x^i,\theta^{old}) \log\left(P(x^i|\theta_k)\right).$$

2. In the M step, we optimize Q wrt π.

$$\pi^{opt}_{z_i=k} = \frac{1}{N}\sum_{i} P(z^i=k|x^i,\theta^{old}). \qquad k=1,2, \quad i=1,2,3.$$

$$\pi_{z_i=1} = \pi_1 = \frac{P(z^1=1|x^1,\theta^{old}) + P(z^2=1|x^2,\theta^{old}) + P(z^3=1|x^3,\theta^{old})}{(N=3)}$$

$$= \frac{1+0.3+0}{3} = \frac{1.3}{3} = \frac{13}{30}.$$

Similarly,

$$\pi_{z_i=2} = \pi_2 = \frac{0+0.7+1}{3} = \frac{17}{30}.$$

Sanity check, $\sum_{k}\pi_k = \frac{13}{30}+\frac{17}{30} = 1.$

3. To derive the M-step for the μ_K's, we look at the parts of Q that depend on μ_K's.

$$f(\mu_K) = \sum_K \sum_i p(z^i = K \mid x^i, \theta^{old}) \log p(x^i \mid \theta_K).$$

$$= -\frac{1}{2} \sum_i p(z^i = K \mid x^i, \theta^{old}) \left[\log|\Sigma_K| + (x^i - \mu_K)^T \Sigma_K^{-1} (x^i - \mu_K)\right].$$

Then,

$$\mu_K^{opt} = \frac{\sum_i p(z^i = K \mid x^i, \theta^{old}) \cdot x^i}{\sum_i p(z^i = K \mid x^i, \theta^{old})}.$$

Therefore,

$$\mu_1 = \frac{p(z^1=1 \mid x^1, \theta^{old}) \cdot x^1 + p(z^2=1 \mid x^2, \theta^{old}) \cdot x^2 + p(z^3=1 \mid x^3, \theta^{old}) \cdot x^3}{p(z^1=1 \mid x^1, \theta^{old}) + p(z^2=1 \mid x^2, \theta^{old}) + p(z^3=1 \mid x^3, \theta^{old})}$$

$$= \frac{1*1 + 0.3*10 + 0*20}{1 + 0.3 + 0}.$$

$$= \frac{40}{13}.$$

Similarly,

$$\mu_2 = \frac{0*1 + 0.7*10 + 1*20}{0 + 0.7 + 1}.$$

$$= \frac{270}{17}.$$

4.

$$\Sigma_k^{opt} = \frac{\sum_i p(z^i = k \mid x^i, \theta^{old})\,(x^i - \mu_k)(x^i - \mu_k)^T}{\sum_i p(z^i = k \mid x^i, \theta^{old})}$$

$$= \frac{\sum_i p(z^i = k \mid x^i, \theta^{old}) . x^i x^{i^T}}{\sum_i p(z^i = k \mid x^i, \theta^{old})} - \mu_k \mu_k^T .$$

Therefore,

$$\Sigma_1 = \frac{1 * \left(1 - \frac{40}{13}\right)^2 + 0.3 * \left(10 - \frac{40}{13}\right)^2 + 0 * \left(20 - \frac{40}{13}\right)^2}{1 + 0.3 + 0} .$$

$$= \frac{(27)^2 + 0.3 * (90)^2 + 0}{1.3 * (13)^2} .$$

$$= 14.3787 .$$

$$\Sigma_2 = \frac{0 * \left(1 - \frac{270}{17}\right)^2 + 0.7 * \left(10 - \frac{270}{17}\right)^2 + 1 * \left(20 - \frac{270}{17}\right)^2}{0 + 0.7 + 1}$$

$$= \frac{\cancel{(253)^2} + 0.7 * (100)^2 + (70)^2}{1.7 * (17)^2}$$

$$= 24.2215 .$$

Thus,

$$\sigma_1 = \sqrt{\Sigma_1} = 3.7919 .$$

$$\sigma_2 = \sqrt{\Sigma_2} = 4.9215 .$$

E-Step.

1. The probability of x^i in cluster c is as following,

$$r_{ik} = \frac{\pi_K \, p(x^i \mid \theta_K^{old})}{\sum_{K'} \pi_{K'} \, p(x^i \mid \theta_{K'}^{old})}.$$

$$r_{11} = \frac{\pi_1 \, p(x^1 \mid \theta_1^{old})}{\pi_1 \, p(x^1 \mid \theta_1^{old}) + \pi_2 \, p(x^1 \mid \theta_2^{old})}. \longrightarrow (a).$$

$$r_{12} = \frac{\pi_2 \, p(x^1 \mid \theta_2^{old})}{\pi_1 \, p(x^1 \mid \theta_1^{old}) + \pi_2 \, p(x^1 \mid \theta_2^{old})}.$$

~~Since,~~

$$\cancel{\pi_1 + \pi_2 = 1.}$$

$$\cancel{r_{12} = \pi_2 = 1 - \pi_1.}$$

Notice that,

$$r_{11} + r_{12} = 1.$$

$$\Rightarrow \quad r_{12} = 1 - r_{11}. \longrightarrow (b).$$

Similarly,

$$r_{21} = \frac{\pi_1 \, p(x^2 \mid \theta_1^{old})}{\pi_1 \, p(x^2 \mid \theta_1^{old}) + \pi_2 \, p(x^2 \mid \theta_2^{old})}. \longrightarrow (c).$$

then, $r_{22} = 1 - r_{21}. \longrightarrow (d).$

Likewise,

$$r_{31} = \frac{\pi_1 \, p(x^3 \mid \theta_1^{old})}{\pi_1 \, p(x^3 \mid \theta_1^{old}) + \pi_2 \, p(x^3 \mid \theta_2^{old})}. \longrightarrow (e)$$

then, $r_{32} = 1 - r_{31}. \longrightarrow (f).$

$$p(x^1|\theta_1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x^1-\mu_1)^2}{\sigma_1^2}\right)$$

$$= \frac{1}{3.7919\sqrt{2\pi}} \cdot \exp\left(-\frac{\left(1-\frac{40}{13}\right)^2}{2\cdot(14.3787)}\right).$$

$$= 0.090554.$$

$$p(x^2|\theta_1) = \frac{1}{3.7919\sqrt{2\pi}} \exp\left(-\frac{\left(10-\frac{40}{13}\right)^2}{2\cdot(14.3787)}\right).$$

$$= 0.019871.$$

$$p(x^3|\theta_1) = \frac{1}{3.7919\sqrt{2\pi}} \exp\left(-\frac{\left(20-\frac{40}{13}\right)^2}{2\cdot(14.3787)}\right).$$

$$= 0.000005.$$

$$p(x^1|\theta_2) = \frac{1}{4.9215\sqrt{2\pi}} \exp\left(-\frac{\left(1-\frac{270}{17}\right)^2}{2\cdot(24.2215)}\right) = 0.000838.$$

$$p(x^2|\theta_2) = \frac{1}{4.9215\sqrt{2\pi}} \exp\left(-\frac{\left(10-\frac{270}{17}\right)^2}{2\cdot(24.2215)}\right) = 0.039682$$

$$p(x^3|\theta_3) = \frac{1}{4.9215\sqrt{2\pi}} \exp\left(-\frac{\left(20-\frac{270}{17}\right)^2}{2\cdot(24.2215)}\right) = 0.057123$$

Therefore,

$$r_{11} = \frac{\frac{13}{30} * 0.090554}{\left(\frac{13}{30} * 0.090554\right) + \left(\frac{17}{30} * 0.000838\right)}$$

$$= \frac{0.0392}{0.0392 + 0.00047487} = 0.9880.$$

$$r_{12} = 1 - r_{11} = 0.0120.$$

$$r_{21} = \frac{\frac{13}{30} * 0.019871}{\left(\frac{13}{30} * 0.019871\right) + \left(\frac{17}{30} * 0.039682\right)} = \frac{0.0086}{0.0311}$$

$$= 0.2765.$$

$$r_{22} = 1 - r_{21} = 0.7235.$$

$$r_{31} = \frac{\frac{13}{30} * 0.000005}{\left(\frac{13}{30} * 0.000005\right) + \left(\frac{17}{30} * 0.057123\right)}$$

$$= \frac{2.1667e-06}{0.0324}$$

$$= 0.00006687.$$

$$r_{32} = 1 - r_{31} = 0.9999.$$

Therefore,

$$R = \begin{bmatrix} 0.9880 & 0.0120 \\ 0.2765 & 0.7235 \\ 0.00006687 & 0.9999 \end{bmatrix}.$$

2.

1. Given,

$$C = \frac{1}{n} X X^T \longrightarrow (1).$$

$$\tilde{X} = (I - v_1 v_1^T) X. \longrightarrow (2).$$

$$\tilde{C} = \frac{1}{n} \tilde{X} \tilde{X}^T \longrightarrow (3).$$

Substituting (2) in equation (3) we get,

$$\tilde{C} = \frac{1}{n} \left\{ (I - v_1 v_1^T) X \right\} \left\{ (I - v_1 v_1^T) X \right\}^T.$$

$$= \frac{1}{n} (I - v_1 v_1^T) X X^T (I^T - v_1 v_1^T).$$

$$= \frac{1}{n} (I - v_1 v_1^T) X X^T (I - v_1 v_1^T).$$

$$= \frac{1}{n} (I - v_1 v_1^T) . (nC) (I - v_1 v_1^T).$$

$$= \frac{1}{n} (nC - n v_1 v_1^T C)(I - v_1 v_1^T).$$

$$= \frac{1}{n} (nC - n v_1 v_1^T C - n C v_1 v_1^T + n v_1 v_1^T C v_1 v_1^T).$$

$$= \frac{1}{n} (X X^T - n v_1 v_1^T \lambda_1 - n C v_1 v_1^T + n v_1 v_1^T \lambda v_1 v_1^T).$$

$$= \frac{1}{n} (X X^T - n \lambda_1 v_1 v_1^T - n \lambda v_1 v_1^T + n \lambda v_1 \underbrace{v_1^T v_1}_{1} v_1^T).$$

$$= \frac{1}{n} (X X^T - n \lambda_1 v_1 v_1^T - n \lambda v_1 v_1^T + n \lambda v_1 v_1^T).$$

$$= \frac{1}{n} (X X^T - n \lambda_1 v_1 v_1^T).$$

$$= \frac{1}{n} X X^T - \lambda v_1 v_1^T.$$

Hence proved.

2.

Since v_j is a principle eigen vector of C, we can write the following,

$$Cv_j = \lambda_j v_j . \longrightarrow (1).$$

where, λ_j is the j^{th} eigenvalue.

If v_j would have been a principle eigen vector of $\tilde{C}$ then we can write the following,

$$\tilde{C} v_j = \frac{1}{n} XX^T v_j - \lambda_1 \cancel{v_1 v_1^T v_j}^{0} . \quad [\text{since, } v_1^T v_j = 0].$$

$$= \frac{1}{n} XX^T v_j .$$

$$= C v_j .$$

$$= \lambda_j v_j . \rightarrow (2).$$

The equation (2) shows that, v_j is an eigenvector of $\tilde{C}$ with λ_j as its corresponding eigenvalue.

3.

The eigen values are numbered in decreasing values of eigenvalues. The order of the first K eigen values is,

$$\lambda_1 > \lambda_2 > \cdots, \lambda_K .$$

If v_2 has to be the first principle eigen vector of $\tilde{C}$ then λ_1 ~~has~~ must be zero. In other words, we need to show $\tilde{C} v_1 = 0$.

$$\tilde{C} v_1 = \frac{1}{n} XX^T v_1 - \lambda_1 v_1 v_1^T v_1$$

$$= C v_1 - \lambda_1 v_1 = \lambda_1 v_1 - \lambda_1 v_1 = 0.$$

Hence proved.

4.

```
function [λ_val, v_vec] = cmpEigs(C, K, f())

    λ_val = zeros(1, K).
    v_vec = zeros(d, K).

    for i = 1:K
        [λ, u] ← f(C);
        C ← C − λuu^T;
        λ_val(i) ← λ;
        v_vec(:, i) ← u;
    end.
```