HW-3 ... ...

1) $f(x|\theta) = \sum_{k=1}^{K} \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right\}, \in \mathbb{R}^D$

where $\theta = \{\pi_1, \mu_1, \Sigma_1\}$

2) Compute $E[\theta|\theta^{(t)}] = E[\{\pi_k, \mu_k, \Sigma_k\}|\theta^{(t)}]$

The π value represents representing the probability that x_i belongs to the kth mixture component

Posterior probability of latent variable Z_i given the observations

$$P(Z_i = k|X_i) = \frac{f(X_i|Z_i=k)\, P(Z_i=k)}{P(X_i)} = \gamma_{Z_i}(k) \quad \text{--} (1)$$

Joint probability of $f(X_i, Z_i|\theta)$

$$f(x_i, z_i|\theta) = f(x_i|z_i)^{I(z_i=k)} f(z_i|\pi)^{I(z_i=k)} \quad f(x_i|z_i)^{I(z_i=k)}$$

$$= \prod_{k=1}^{K} \pi_k^{I(z_i=k)} \mathcal{N}(x_i|\mu_k, \Sigma_k)^{I(z_i=k)}$$

Likelihood of complete data $\{X_i, Z_i\}_{i=1}^n$

$$L(\theta|X, Z) = \prod_{i=1}^{n} f(x_i, z_i|\theta)$$

$$= \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{I(z_i=k)} \mathcal{N}(x_i|\mu_k, \Sigma_k)^{I(z_i=k)}$$

Log-likelihood of complete data $\{X_i, Z_i\}_{i=1}^n$

$$\ell(\theta|X, Z) = \log L(\theta|X, Z)$$

$$= \log \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{I(z_i=k)} \mathcal{N}(x_i|\mu_k, \Sigma_k)^{I(z_i=k)}$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} I(z_i=k)\left[\log \pi_k + \log \mathcal{N}(x_i|\mu_k, \Sigma_k)\right]$$

Taking expectation of the complete log likelihood with the posterior of the latent variable.

$$E[\ell(\theta|X, Z)|X, \theta^{(t)}] = \underset{Z \sim f(Z|X, \theta^{(t)})}{E}\left[\ell(\theta|X, Z)\right]$$

$$= Q(\theta|\theta^{(t)})$$

$$Q(\theta|\theta^{(t)}) = \underset{Z|X}{E}\left[\ell(\theta|X, Z)\right]$$

$$= \underset{Z|X}{E}\left[\sum_{i=1}^{n} \sum_{k=1}^{K} I(z_i=k)\left[\log \pi_k + \log N(x_i|\mu_k, \Sigma_k)\right]\right]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \underset{Z|X}{E}\left[I(z_i=k)\right]\left[\log \pi_k + \log N(x_i|\mu_k, \Sigma_k)\right]$$

Since $\underset{Z|X}{E}\left[I(z_i=k)\right] = P(z_i=k|x)$, from (1) we have,

$$P(z_i=k|x) = \gamma_{z_i}(k)$$

$$\boxed{\therefore Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{z_i}(k)\left[\log \pi_k + \log N(x_i|\mu_k, \Sigma_k)\right]}$$

where $\log N(x_i|\mu_k, \Sigma_k) = \log\left[\frac{1}{|\det(2\pi\Sigma_k)|} \cdot \exp\left[-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right]\right]$

$$= -\frac{1}{2}\log \det(2\pi\Sigma_k) - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)$$

3) In M-step, we will maximize $Q(\theta|\theta^{(t)})$ (in (2))

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{z_i}(k)\left[\log \pi_k + \log N(x_i|\mu_k, \Sigma_k)\right] \text{--} (1)$$

where $\gamma_{z_i}(k) = \frac{f(X_i|Z_i=k)\, P(Z_i=k)}{P(X_i)}$

subject to $\sum_{k=1}^{K} \pi_k = 1$

Because of equality constraint of that the objective is constraint one can find optimization problem can be solved using Lagrange's multiplier.

$$L(\theta, \lambda) = \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{z_i}(k)\left[\log \pi_k + \log N(x_i|\mu_k, \Sigma_k)\right] + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)$$

3) Starting with $\pi_1, \pi_2 \ldots \pi_K$.

For some k,

$$\frac{\partial L(\theta, \lambda)}{\partial \pi_k} = \frac{1}{\pi_k} \sum_{i=1}^{n} \gamma_{z_i}(k) + \lambda \quad (2)$$

Setting it to 0 & solving for π_k.

$$\pi_k = -\frac{1}{\lambda} \sum_{i=1}^{n} \gamma_{z_i}(k)$$

Plugging this into the constraint $\sum_{k=1}^{K} \pi_k = 1$

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} -\frac{1}{\lambda} \sum_{i=1}^{n} \gamma_{z_i}(k) = 1$$

$$\lambda = -\sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{z_i}(k) \qquad \left(\because \sum_{k=1}^{K} \gamma_{z_i}(k) = 1\right)$$

$$\lambda = -n$$

Plugging $\lambda = -n$ in (2)

$$\frac{1}{\pi_k} \sum_{i=1}^{n} \gamma_{z_i}(k) - n = 0$$

$$\Rightarrow \boxed{\pi_k = \frac{1}{n} \sum_{i=1}^{n} \gamma_{z_i}(k)}$$

II) Now estimating μ_k

For some k,

$$L(\theta, \lambda) = \sum_{i=1}^{n} \gamma_{z_i}(k) \log \mathcal{N}(x_i|\mu_k, \Sigma_k) + \text{terms not containing } \mu_k, \Sigma_k$$

$$= \sum_{i=1}^{n} \gamma_{z_i}(k)\left[\log \frac{1}{\sqrt{\det(2\pi\Sigma_k)}} \cdot \exp\left(-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right)\right]$$

$$= \sum_{i=1}^{n} \gamma_{z_i}(k)\left[\frac{1}{2}\log(\det(\Sigma_k)) - \frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right]$$

$$= -\frac{1}{2}\left[\sum_{i=1}^{n} \gamma_{z_i}(k) \log(\det(\Sigma_k))\right] + \sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k) \quad \text{--} (4)$$

3) Taking derivative of (4) wrt μ_k & setting to 0.

$$-\frac{1}{2}\frac{\partial\left[\sum_{i=1}^{n} \gamma_{z_i}(k) \log(\det(\Sigma_k))\right] + \sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)}{\partial \mu_k} = 0$$

$$-\frac{1}{2}\frac{\partial \sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)}{\partial \mu_k} = 0$$

$$\sum_{i=1}^{n} \gamma_{z_i}(k) \Sigma_k^{-1}(\mu_k - x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \gamma_{z_i}(k) \Sigma_k^{-1} \mu_k = \sum_{i=1}^{n} \gamma_{z_i}(k) \Sigma_k^{-1} x_i$$

$$\Rightarrow \mu_k \cancel{\Sigma_k^{-1}} \sum_{i=1}^{n} \gamma_{z_i}(k) = \cancel{\Sigma_k^{-1}} \sum_{i=1}^{n} \gamma_{z_i}(k) x_i$$

$$\boxed{\mu_k = \frac{\sum_{i=1}^{n} \gamma_{z_i}(k) x_i}{\sum_{i=1}^{n} \gamma_{z_i}(k)}}$$

3) Taking derivative of (4) wrt Σ_k^{-1} & setting it to 0

$$-\frac{1}{2}\frac{\partial\left[\sum_{i=1}^{n} \gamma_{z_i}(k) \log(\det(\Sigma_k)) + \sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right]}{\partial \Sigma_k^{-1}} = 0$$

$$\frac{1}{2}\frac{\partial \sum_{i=1}^{n} \gamma_{z_i}(k) \log(\det(\Sigma_k^{-1}))}{\partial \Sigma_k^{-1}} - \frac{1}{2}\frac{\partial \sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)}{\partial \Sigma_k^{-1}} = 0$$

$$\frac{1}{2}\sum_{i=1}^{n} \gamma_{z_i}(k) \Sigma_k - \frac{1}{2}\sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)(x_i-\mu_k)^T = 0$$

$$\Sigma_k \sum_{i=1}^{n} \gamma_{z_i}(k) = \sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)(x_i-\mu_k)^T$$

$$\boxed{\Sigma_k = \frac{\sum_{i=1}^{n} \gamma_{z_i}(k)(x_i-\mu_k)(x_i-\mu_k)^T}{\sum_{i=1}^{n} \gamma_{z_i}(k)}}$$

2) (a) $X = (x_1, x_2, x_3)$: 3 categories

$$f(x|\theta) = \frac{(x_1+x_2+x_3)!}{x_1!\, x_2!\, x_3!}\left(\frac{2+\theta}{6}\right)^{x_1}\left(\frac{1-\theta}{3}\right)^{x_2}\left(\frac{\theta}{6}\right)^{x_3}$$

Let complete data be $Y = (y_1, y_2, y_3, y_4)$

$$\Rightarrow g_\theta = \left(\underbrace{\frac{2}{6}, \frac{\theta}{6}}_{x_1}, \underbrace{\frac{1-\theta}{3}}_{x_2}, \underbrace{\frac{\theta}{6}}_{x_3}\right)$$

So,

$$f(y|\theta) = \frac{(y_1+y_2+y_3+y_4)!}{y_1!\, y_2!\, y_3!\, y_4!}\left(\frac{2}{6}\right)^{y_1}\left(\frac{\theta}{6}\right)^{y_2}\left(\frac{1-\theta}{3}\right)^{y_3}\left(\frac{\theta}{6}\right)^{y_4} \quad \text{--} (1)$$

E-step: Estimating sufficient statistic for complete data Y given observed data X.

We know that,

$$(y_3, y_4) = (x_2, x_3)$$

$$y_1^t = x_1 \cdot \frac{2/6}{2/6 + \theta/6} = \frac{2x_1}{2+\theta^t} \qquad [x_1 = y_1 + y_2]$$

$$y_2^t = x_1 \frac{\theta^t/6}{\left(\frac{2+\theta^t}{6}\right)} = \frac{\theta^t x_1}{2+\theta^t} \quad (5)$$

$$Q(\theta|\theta^t) = E_{Y|X,\theta^t}\left[\log f(y|\theta)\right]$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\ E_{Y|X,\theta^t}\left[\log f(y|\theta)\right]$$

$$\sum_{i=1}^{n} \gamma_{z_i} = 1$$

$$\lambda = -n$$

Since we are maximizing θ, we are only interested in terms related to θ.

So, taking log of eq (1),

$$Q(\theta|\theta^t) = E\left[\log\left[\left(\frac{\theta}{6}\right)^{y_2+y_4}\left(\frac{1-\theta}{3}\right)^{y_3}\right]\right]$$

$$\Rightarrow E\left[(y_2+y_4)\log\frac{\theta}{6} + y_3 \log\frac{(1-\theta)}{3}\right]$$

Using eq (5)

$$Q(\theta|\theta^t) = E\left[\left(\frac{\theta^t x_1}{2+\theta^t} + x_3\right)\log\theta + x_2 \log(1-\theta)\right]$$

M-Step

Maximizing $Q(\theta|\theta^t)$, diff wrt $\partial\theta$ & setting it to 0

$$\frac{\partial Q(\theta|\theta^t)}{\partial\theta} = \left(\frac{x_1\theta^t}{2+\theta^t} + x_3\right)\frac{1}{\theta} + \frac{x_2(-1)}{1-\theta} = 0$$

$$(1-\theta)\left[\frac{x_1\theta^t}{2+\theta^t} + x_3\right] = \theta x_2$$

$$\boxed{\hat{\theta} = \left(\frac{x_1\theta^t}{2+\theta^t} + x_3\right) \cdot \frac{1}{x_2 + x_3 + \left(\frac{x_1\theta^t}{2+\theta^t}\right)}}$$