Gaussian Mixture Model (GMM)

Here we set following symbols

$X : l \times N$ (n= 1 ~ N)

$$X = [\boldsymbol{x_1}\boldsymbol{x_2}\ldots\boldsymbol{x_n}\ldots\boldsymbol{x_N}]$$

$\pi_k : 1 \times 1$ (k = 1~K)

$\mu_k : l \times 1$ (k = 1~K)

Σ_k: $l \times l$ (k = 1~K)

$\boldsymbol{\phi}$: $l \times N$

$$\boldsymbol{\phi} = [\phi_k(\boldsymbol{x_n})] = [\cdots \boldsymbol{\phi}(\boldsymbol{x_n}) \cdots \boldsymbol{\phi}(\boldsymbol{x_N})] \ = \begin{bmatrix} \vdots \\ \boldsymbol{\phi_k}(\boldsymbol{X})^T \\ \vdots \end{bmatrix}$$

$\phi_k(\boldsymbol{x})$:

$$\phi_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{l/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu_k)^T\Sigma_k^{-1}(\boldsymbol{x}-\mu_k)\right)$$

$$\ln\phi_k(\boldsymbol{x}) = -\frac{l}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(\boldsymbol{x}-\mu_k)^T\Sigma_k^{-1}(\boldsymbol{x}-\mu_k)$$

$\boldsymbol{Z}$: $K \times N$

$$\begin{aligned}\boldsymbol{Z} &= [\ldots\boldsymbol{z_n}\ldots\boldsymbol{z_N}] \\ &= \begin{bmatrix} \vdots \\ \boldsymbol{z_k}^T \\ \vdots \end{bmatrix}\end{aligned}$$

z_{kn}: 1×1 (k = 1 ~ K)

$\boldsymbol{z_k} = [0, 1, \ldots, 1, \ldots, 0]^T$ represents probability K states in N samples ($N \times 1$)

$\boldsymbol{z_n} = [0, 0, \ldots, 1, \ldots, 0]^T$ represents probability of length = K states, ($K \times 1$)only when $z_k = 1$, it would be counted into p, **random variable**

$$p(\boldsymbol{z}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{1^\top z_k}$$

$$p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \prod_{k=1}^{K} \phi_k(\boldsymbol{x})^{1^\top z_k}$$

Goal:

$$\underset{\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}}{\arg\max} \quad p(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$$

think about the Expectation–maximization algorithm

$$\underset{\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}}{\arg\max} \quad \ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q\|p)$$

where

$$\begin{aligned}
\mathcal{L}(q, \boldsymbol{\theta}) &= \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\} \\
\mathrm{KL}(q\|p) &= -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\}
\end{aligned}$$

suppose $\boldsymbol{z_1}\boldsymbol{z_2}\ldots\boldsymbol{z_n}\ldots\boldsymbol{z_N}$ independent

Here $p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})$

$$\begin{aligned}
p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) &= p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) p(\boldsymbol{Z}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) \\
&= p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) p(\boldsymbol{Z}|\boldsymbol{\pi}) \\
&= \prod_{n=1}^{N} p(\boldsymbol{x_n}|\boldsymbol{Z}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) \prod_{n=1}^{N} p(\boldsymbol{z_n}|\boldsymbol{\pi}) \\
&= \prod_{n=1}^{N} p(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) \prod_{n=1}^{N} p(\boldsymbol{z_n}|\boldsymbol{\pi})
\end{aligned}$$

Hence

$$
\begin{aligned}
\ln\Big(p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})\Big) &= \sum_{n=1}^{N}\ln p\left(\boldsymbol{x_n}|\boldsymbol{z_n},\boldsymbol{\Sigma},\boldsymbol{\mu}\right)+\sum_{n=1}^{N}\ln p\left(\boldsymbol{z_n}|\boldsymbol{\pi}\right)\\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}(\ln\pi_k)z_{kn}+\sum_{n=1}^{N}\sum_{k=1}^{K}(\ln\phi_k(\boldsymbol{x_n}))z_{kn}\\
&= \ln\boldsymbol{\pi}^T\cdot\boldsymbol{Z}\cdot ones(N,1)+\\
&\sum_{n=1}^{N}\sum_{k=1}^{K}\Big[-\frac{l}{2}\ln(2\pi)-\frac{1}{2}\ln|\Sigma_k|-\frac{1}{2}(\boldsymbol{x_n}-\mu_k)^T\Sigma_k^{-1}(\boldsymbol{x_n}-\mu_k)\Big]z_{kn}\\
&= (\ln\boldsymbol{\pi}^T-\frac{1}{2}\ln|\boldsymbol{\Sigma}|^T)\cdot\boldsymbol{Z}\cdot ones(N,1)-\frac{l}{2}\ln(2\pi)ones(1,K)\cdot\boldsymbol{Z}\cdot ones(N,\\
&-\frac{1}{2}\sum_{k=1}^{K}\Big[\mathrm{tr}\Big((X-\mu_k\cdot 1^T)^T\Sigma_k^{-1}(X-\mu_k\cdot 1^T)\cdot\mathrm{diag}(\boldsymbol{z_k})\Big)\Big]\\
&= (\ln\boldsymbol{\pi}^T-\frac{1}{2}\ln|\boldsymbol{\Sigma}|^T)\cdot\boldsymbol{Z}\cdot ones(N,1)-\frac{l}{2}\ln(2\pi)N\\
&-\frac{1}{2}\sum_{k=1}^{K}\Big[\mathrm{tr}\Big((X-\mu_k\cdot 1^T)^T\Sigma_k^{-1}(X-\mu_k\cdot 1^T)\cdot\mathrm{diag}(\boldsymbol{z_k})\Big)\Big]
\end{aligned}
$$

notice: X is fixed, Z,θ are variables, we turn to replace it with $q(Z),\theta$

EM method

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta})=\mathcal{L}(q,\boldsymbol{\theta})+\mathrm{KL}(q|p)$$

[M step] fix $q(Z)$, change θ

fix $q\equiv q^k=p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^k\right)=p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)$

$$
\begin{aligned}
\theta^{k+1} &\equiv \underset{\theta}{\operatorname{argmax}}\mathcal{L}(q^k,\boldsymbol{\theta})\\
&= \underset{\theta}{\operatorname{argmax}}\sum_{\boldsymbol{Z}}q^k\ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{q^k}\right\}\\
&= \underset{\theta}{\operatorname{argmax}}\sum_{\boldsymbol{Z}}q^k\ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})-\sum_{\boldsymbol{Z}}q^k\ln(q^k)\\
&= \underset{\theta}{\operatorname{argmax}}\sum_{\boldsymbol{Z}}q^k\ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})\\
&= \underset{\theta}{\operatorname{argmax}}\boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{ old}}\right)
\end{aligned}
$$

$$\mathcal{L}(q^k,\boldsymbol{\theta}^{k+1})>\mathcal{L}(q^k,\boldsymbol{\theta}^k)=\ln p(\boldsymbol{X}|\boldsymbol{\theta}^k)$$

where $\boldsymbol{\theta}^{\text{ old}}\equiv\boldsymbol{\theta}^k$

$$\boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) \equiv \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})$$

[E step] fix θ, change $q(Z)$

fix $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^{k+1}$

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) = \mathcal{L}(q, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q|p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1}\right)) = \text{const}$$

when $\mathrm{KL}(q|p) \downarrow$, wen have $\mathcal{L}(q, \boldsymbol{\theta}^{\boldsymbol{k+1}}) \uparrow$

When $\mathrm{KL}(q|p) \downarrow$ reaches the minimal value 0，then q have $q^{k+1} = p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta})$

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) == \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1})$$

update $\boldsymbol{\theta}^{\text{old}}$ ， to keep current $\boldsymbol{\theta}$ value

$$\boldsymbol{\theta}^{\text{old}} \Leftarrow \boldsymbol{\theta}^{\boldsymbol{k+1}}$$

$$\begin{aligned} q^{k+1} &\equiv \underset{q}{\operatorname{argmax}} \mathcal{L}(q, \boldsymbol{\theta}^{k+1}) \\ &\equiv \underset{q}{\operatorname{argmin}} \mathrm{KL}(q|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) = 0 \\ &= p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1}) \end{aligned}$$

$$\begin{aligned} \ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) &\equiv \mathcal{L}(q, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) \\ &= \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q^k|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) \\ &= \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q^{k+1}|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) \\ &= \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) \end{aligned}$$

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) = \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) > \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1})$$

in all

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) = \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) > \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1}) > \ln p(\boldsymbol{X}|\boldsymbol{\theta}^k) = \mathcal{L}(q^k, \boldsymbol{\theta}^k)$$

So

$$
\begin{aligned}
\lim_{k\to\infty} \ln p(\boldsymbol{X}|\boldsymbol{\theta}^k) = \lim_{k\to\infty} \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1}) = \max_\theta \ln p(\boldsymbol{X}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}_{max} &\equiv \max_\theta \ln p(\boldsymbol{X}|\boldsymbol{\theta}) \\
&= \lim_{k\to\infty} \theta^{k+1} \\
&= \lim_{k\to\infty} \operatorname*{argmax}_\theta \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) \\
&= \lim_{k\to\infty} \operatorname*{argmax}_\theta \boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^k\right)
\end{aligned}
$$

Calculation of [M step] fix $q(Z)$, change θ

fix $q \equiv q^k = p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^k\right) = p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right)$

where $\boldsymbol{\theta}^{\text{ old}} \equiv \boldsymbol{\theta}^k$

$$
\boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old}}\right) \equiv \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})
$$

Here

$$
\begin{aligned}
\theta^{k+1} &\equiv \operatorname*{argmax}_\theta \mathcal{L}(q^k, \boldsymbol{\theta}) \\
&= \operatorname*{argmax}_\theta \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) - \sum_{\boldsymbol{Z}} q^k \ln(q^k) \\
&= \operatorname*{argmax}_\theta \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) \\
&= \operatorname*{argmax}_\theta \boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old}}\right)
\end{aligned}
$$

Now find close form of Q

$\sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right)$ is [probability distribution] random variable **of** Z [probability distribution] random variable

Here $q^k = p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^k\right) = p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right)$

$$
\begin{aligned}
\boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old}}\right) &\equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z} \mid \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right) \ln p(\boldsymbol{X}, \boldsymbol{Z} \mid \boldsymbol{\theta}) \\
&= \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z} \mid \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right) \Bigg\{ (\ln \boldsymbol{\pi}^T - \frac{1}{2}|\boldsymbol{\Sigma}|^T) \cdot \boldsymbol{Z} \cdot ones(N,1) \\
&- \frac{l}{2} \ln(2\pi) ones(1,K) \cdot \boldsymbol{Z} \cdot ones(N,1) \\
&- \frac{1}{2} \sum_{k=1}^{K} \left[\operatorname{tr}\Big((X - \mu_k \cdot 1^T)^T \Sigma_k^{-1} (X - \mu_k \cdot 1^T) \cdot \operatorname{diag}(\boldsymbol{z_k}) \Big) \right] \Bigg\} \\
&= \left\{ \left[\sum_{n=1}^{N} \sum_{z_n} p\left(\boldsymbol{z_n} \mid \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old}}\right) \boldsymbol{z_n}^T \right] (\ln \boldsymbol{\pi} - \frac{1}{2}|\boldsymbol{\Sigma}| - \frac{l}{2} \ln(2\pi) \cdot 1.) \right\} \\
&- \frac{1}{2} \left\{ \sum_{n=1}^{N} \sum_{k=1}^{K} \left[\sum_{z_n} p\left(\boldsymbol{z_n} \mid \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old}}\right) (\boldsymbol{x_n} - \mu_k)^T \Sigma_k^{-1} (\boldsymbol{x_n} - \mu_k) z_{kn} \right] \right\} \\
&= \left[\boldsymbol{\gamma} \cdot 1.\right]^T (\ln \boldsymbol{\pi} - \frac{1}{2}|\boldsymbol{\Sigma}| - \frac{l}{2} \ln(2\pi) \cdot 1.) \\
&- \frac{1}{2} \sum_{k=1}^{K} \operatorname{tr}\Big(diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T \Sigma_k^{-1} (X - \mu_k \cdot 1^T) \Big)
\end{aligned}
$$

Here

$$
p(\boldsymbol{z_n} \mid \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{old}}) = \frac{p(\boldsymbol{x_n} \mid \boldsymbol{z_n}, \boldsymbol{\theta}^{\text{old}}) p(\boldsymbol{z_n} \mid \boldsymbol{\theta}^{\text{old}})}{\sum\limits_{z_n} p(\boldsymbol{x_n} \mid \boldsymbol{z_n}, \boldsymbol{\theta}^{\text{old}}) p(\boldsymbol{z_n} \mid \boldsymbol{\theta}^{\text{old}})}
$$

So, have $1.^T \boldsymbol{\gamma} = ones(1, N)$

$$
p(\boldsymbol{z_n} = k \mid \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{old}}) = \frac{\phi_k^{\text{old}}(\boldsymbol{x_n}) \pi_k^{\text{old}}}{\sum\limits_{k=1}^{K} \phi_k^{\text{old}}(\boldsymbol{x_n}) \pi_k^{\text{old}}} \equiv \gamma_{kn}
$$

$$
\boldsymbol{\gamma} = \boldsymbol{\phi}^{\text{old}}(X) \odot (\pi_k^{\text{old}} \cdot 1.^T) \odot \big(\frac{1.}{1. \cdot (\boldsymbol{\pi}^{\text{old}})^T \boldsymbol{\phi}^{\text{old}}(X)}\big)
$$

maximize Q

$$
\begin{aligned}
\max_{\boldsymbol{\theta}} \quad \boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old}}\right) &\equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z} \mid \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right) \ln p(\boldsymbol{X}, \boldsymbol{Z} \mid \boldsymbol{\theta}) \\
&= \left[\boldsymbol{\gamma} \cdot 1.\right]^T (\ln \boldsymbol{\pi} - \frac{1}{2}|\boldsymbol{\Sigma}| - \frac{l}{2} \ln(2\pi) \cdot 1.) \qquad \text{s.t.} \quad 1.^T \cdot \boldsymbol{\pi} = \\
&- \frac{1}{2} \sum_{k=1}^{K} \operatorname{tr}\Big(diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T \Sigma_k^{-1} (X - \mu_k \cdot 1^T) \Big)
\end{aligned}
$$

define $L \equiv Q - \lambda(1.^T \cdot \boldsymbol{\pi} - 1)$

for $\boldsymbol{\pi}$

$$\frac{\partial L}{\partial \boldsymbol{\pi}} = \big[\boldsymbol{\gamma} \cdot 1.\big] \odot \frac{1.}{\boldsymbol{\pi}} - \lambda 1. = 0$$

so

$$\big[\boldsymbol{\gamma} \cdot 1.\big] = \lambda \boldsymbol{\pi}$$

Hence

$$\begin{aligned}
\lambda &= 1.^T\big[\boldsymbol{\gamma} \cdot 1.\big] = [1.^T\boldsymbol{\gamma}] \cdot 1. = ones(1,N) \cdot 1. = N \\
\boldsymbol{\pi} &= \frac{\big[\boldsymbol{\gamma} \cdot 1.\big]}{1.^T\big[\boldsymbol{\gamma} \cdot 1.\big]} = \frac{\big[\boldsymbol{\gamma} \cdot 1.\big]}{N}
\end{aligned}$$

for μ_k

$$\frac{\partial L}{\partial \mu_k} = -\frac{1}{2} \cdot 2 \cdot (-1)\Sigma_k^{-1}(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot 1. = 0$$

so

$$\begin{gathered}
(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot 1. = 0 \\
(X - \mu_k \cdot 1^T)\boldsymbol{\gamma}_k = 0
\end{gathered}$$

Hence

$$\mu_k = \frac{X\boldsymbol{\gamma}_k}{1.^T\boldsymbol{\gamma}_k}$$

for Σ_k

$$\begin{aligned}
dL &= \big[\boldsymbol{\gamma}_k^T \cdot 1.\big]^T(\frac{-1}{2})\Sigma_k^{-1}d\Sigma_k \\
&\quad - \frac{1}{2}\mathrm{tr}\Big(diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T d\Sigma_k^{-1}(X - \mu_k \cdot 1^T)\Big) \\
&= \big[\boldsymbol{\gamma}_k^T \cdot 1.\big]^T(\frac{-1}{2})\Sigma_k^{-1}d\Sigma_k \\
&\quad - \frac{1}{2}\mathrm{tr}\Big((X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T d\Sigma_k^{-1}\Big) \\
&= -(\frac{1}{2})\big[\boldsymbol{\gamma}_k^T \cdot 1.\big]^T\Sigma_k^{-1}d\Sigma_k \\
&\quad + \frac{1}{2}\mathrm{tr}\Big((X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T\Sigma_k^{-1}d\Sigma_k\Sigma_k^{-1}\Big) \\
&= -(\frac{1}{2})\big[\boldsymbol{\gamma}_k^T \cdot 1.\big]^T\Sigma_k^{-1}d\Sigma_k \\
&\quad + \frac{1}{2}\mathrm{tr}\Big(\Sigma_k^{-1}(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T\Sigma_k^{-1}d\Sigma_k\Big)
\end{aligned}$$

note:$0 = d(\Sigma_k^{-1}\Sigma_k) = (d\Sigma_k^{-1})\Sigma + \Sigma_k^{-1}d\Sigma_k$, so $d\Sigma_k^{-1} = -\Sigma_k^{-1}d\Sigma_k\Sigma_k^{-1}$

So

$$\frac{\partial L}{\partial \Sigma_k} = -(\frac{1}{2})\left[\boldsymbol{\gamma}_k^T \cdot 1.\right]\Sigma_k^{-T} \\ + \frac{1}{2}\left(\Sigma_k^{-T}(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T \Sigma_k^{-T}\right) \\ = 0$$

So

$$\left[\boldsymbol{\gamma}_k^T \cdot 1.\right] I_{l\times l} = \Sigma_k^{-T}(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T$$
$$\left[\boldsymbol{\gamma}_k^T \cdot 1.\right] I_{l\times l} = (X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T \Sigma_k^{-1}$$
$$\Sigma_k = \frac{(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T}{\left[\boldsymbol{\gamma}_k^T \cdot 1.\right]}$$

To sum up

For M step

$$\boldsymbol{\pi} = \frac{\left[\boldsymbol{\gamma} \cdot 1.\right]}{1.^T\left[\boldsymbol{\gamma} \cdot 1.\right]} = \frac{\left[\boldsymbol{\gamma} \cdot 1.\right]}{N}$$
$$\mu_k = \frac{X\boldsymbol{\gamma}_k}{1.^T\boldsymbol{\gamma}_k}$$
$$\Sigma_k = \frac{(X - \mu_k \cdot 1^T)diag(\boldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T}{1.^T\boldsymbol{\gamma}_k} = \frac{X diag(\gamma_k^\top)X^T}{1.^T\boldsymbol{\gamma}_k} - \mu_k\mu_k^\top$$

where, have $1.^T\boldsymbol{\gamma} = ones(1, N)$

For E step

$$p(\boldsymbol{z_n} = k|\boldsymbol{x_n}, \boldsymbol{\theta}^{\text{old}}) = \frac{\phi_k^{\text{old}}(\boldsymbol{x_n})\pi_k^{\text{old}}}{\sum\limits_{k=1}^{K}\phi_k^{\text{old}}(\boldsymbol{x_n})\pi_k^{\text{old}}} \equiv \gamma_{kn}$$
$$\boldsymbol{\gamma} = \boldsymbol{\phi}^{\text{old}}(X) \odot (\pi^{\text{old}} \cdot 1.^T) \odot (\frac{1.}{1. \cdot (\boldsymbol{\pi}^{\text{old}})^T\boldsymbol{\phi}^{\text{old}}(X)})$$

here $\boldsymbol{\gamma}_k$ is (N x 1) matrix

$$\boldsymbol{\gamma} = \begin{bmatrix} \vdots \\ \boldsymbol{\gamma}_k^T \\ \vdots \end{bmatrix}$$