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

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

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

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

Σ_k: $l \times l$ (k = 1~K)

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

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

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

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

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

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

z_{kn}: 1×1 (k = 1 ~ K)

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

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

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

Goal:

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

期望最大化算法，或者**EM**算法

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

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

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

Hence

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

notice: X is fixed, Z, θ are variables,转而用$q(Z), \theta$ 替代；

EM法

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

[M步骤]fix $q(Z)$, change θ

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

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

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步骤]fix θ, change $q(Z)$

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

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

第2部分$\mathrm{KL}(q|p)\downarrow$,则$\mathcal{L}(q, \boldsymbol{\theta}^{\boldsymbol{k+1}})\uparrow$

而$\mathrm{KL}(q|p)\downarrow$取min值0，此时$q$取$q^{k+1} = p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta})$

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

更新$\boldsymbol{\theta}^{\text{ old }}$，保留此时的$\boldsymbol{\theta}$值

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

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

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

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

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

Here

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

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

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

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

maximize Q

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

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

for $\boldsymbol{\pi}$

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

so

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

Hence

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

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

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

Hence

$$\mu_k = \frac{X\gamma_k}{1.^T\gamma_k}$$

for Σ_k

$$\begin{aligned}
dL &= \left[\gamma_k^T \cdot 1.\right]^T (\frac{-1}{2})\Sigma_k^{-1} d\Sigma_k \\
&- \frac{1}{2}\operatorname{tr}\Big(diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T d\Sigma_k^{-1}(X-\mu_k\cdot 1^T)\Big) \\
&= \left[\gamma_k^T \cdot 1.\right]^T (\frac{-1}{2})\Sigma_k^{-1} d\Sigma_k \\
&- \frac{1}{2}\operatorname{tr}\Big((X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T d\Sigma_k^{-1}\Big) \\
&= -(\frac{1}{2})\left[\gamma_k^T \cdot 1.\right]^T \Sigma_k^{-1} d\Sigma_k \\
&+ \frac{1}{2}\operatorname{tr}\Big((X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T \Sigma_k^{-1} d\Sigma_k \Sigma_k^{-1}\Big) \\
&= -(\frac{1}{2})\left[\gamma_k^T \cdot 1.\right]^T \Sigma_k^{-1} d\Sigma_k \\
&+ \frac{1}{2}\operatorname{tr}\Big(\Sigma_k^{-1}(X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T \Sigma_k^{-1} d\Sigma_k\Big)
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial L}{\partial \Sigma_k} &= -(\frac{1}{2})\left[\gamma_k^T \cdot 1.\right]\Sigma_k^{-T} \\
&+ \frac{1}{2}\Big(\Sigma_k^{-T}(X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T \Sigma_k^{-T}\Big) \\
&= 0
\end{aligned}$$

So

$$\begin{aligned}
\left[\gamma_k^T \cdot 1.\right] I_{l\times l} &= \Sigma_k^{-T}(X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T \\
\left[\gamma_k^T \cdot 1.\right] I_{l\times l} &= (X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T \Sigma_k^{-1} \\
\Sigma_k &= \frac{(X-\mu_k\cdot 1^T)diag(\gamma_k^T)\cdot(X-\mu_k\cdot 1^T)^T}{\left[\gamma_k^T \cdot 1.\right]}
\end{aligned}$$

To sum up

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

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

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

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

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

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