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

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

 $\pi_k: \mathbf{1} \times \mathbf{1} \ (\mathsf{k} = 1 \sim \mathsf{K})$

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

 Σ_k : $l \times l$ (k = 1~K)

 ϕ : $l \times N$

$$oldsymbol{\phi} = [\phi_k(oldsymbol{x_n})] = [\cdots \phi(oldsymbol{x_n}) \cdots \phi(oldsymbol{x_N})] = egin{bmatrix} dots \ \phi_k(oldsymbol{X})^T \ dots \ \end{matrix}$$

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

$$egin{aligned} \phi_k(m{x}) &= rac{1}{(2\pi)^{l/2} |\Sigma_k|^{1/2}} \exp(-rac{1}{2} (m{x} - \mu_k)^T \Sigma_k^{-1} (m{x} - \mu_k)) \ \ln \phi_k(m{x}) &= -rac{l}{2} \ln (2\pi) - rac{1}{2} \ln |\Sigma_k| - rac{1}{2} (m{x} - \mu_k)^T \Sigma_k^{-1} (m{x} - \mu_k) \end{aligned}$$

 $Z: K \times N$

$$oldsymbol{Z} = [\dots oldsymbol{z_n} \dots oldsymbol{z_N}] \ = egin{bmatrix} dots \ oldsymbol{z_k}^T \ oldsymbol{z_k}^T \ dots \end{bmatrix}$$

 $z_{kn}: 1 \times 1 \text{ (k = 1 ~ K)}$

 $oldsymbol{z_k} = [0,1,\ldots,1,\ldots,0]^T$ represent probability K states in N samples (N imes 1)

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

$$egin{aligned} p\left(oldsymbol{z} \middle| oldsymbol{\pi}
ight) &= \prod_{k=1}^K \pi_k^{z_k} \ p\left(oldsymbol{x} \middle| oldsymbol{z}, oldsymbol{\Sigma}, oldsymbol{\mu}
ight) &= \prod_{k=1}^K \phi_k(oldsymbol{x})^{z_k} \end{aligned}$$

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

$$egin{aligned} \mathcal{L}(q,m{ heta}) &= \sum_{m{Z}} q(m{Z}) \ln \Big\{ rac{p(m{X},m{Z}|m{ heta})}{q(m{Z})} \Big\} \ \mathrm{KL}(q\|p) &= -\sum_{m{Z}} q(m{Z}) \ln \Big\{ rac{p(m{Z}|m{X},m{ heta})}{q(m{Z})} \Big\} \end{aligned}$$

suppose $z_1 z_2 \dots z_n \dots z_N$ independent

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

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

Hence

$$\begin{split} \ln\left(p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})\right) &= \sum_{n=1}^{N} \ln p\left(\boldsymbol{x}_{n}|\boldsymbol{z}_{n},\boldsymbol{\Sigma},\boldsymbol{\mu}\right) + \sum_{n=1}^{N} \ln p\left(\boldsymbol{z}_{n}|\boldsymbol{\pi}\right) \\ &= \sum_{n=1}^{N} \sum_{k=1}^{K} (\ln \pi_{k}) z_{kn} + \sum_{n=1}^{N} \sum_{k=1}^{K} (\ln \phi_{k}(\boldsymbol{x}_{n})) z_{kn} \\ &= \ln \boldsymbol{\pi}^{T} \cdot \boldsymbol{Z} \cdot ones(N,1) + \\ &\sum_{n=1}^{N} \sum_{k=1}^{K} \left[-\frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_{k}| - \frac{1}{2} (\boldsymbol{x}_{n} - \mu_{k})^{T} \boldsymbol{\Sigma}_{k}^{-1} (\boldsymbol{x}_{n} - \mu_{k}) \right] z_{kn} \\ &= (\ln \boldsymbol{\pi}^{T} - \frac{1}{2} |\boldsymbol{\Sigma}|^{T}) \cdot \boldsymbol{Z} \cdot ones(N,1) - \frac{l}{2} \ln(2\pi) ones(1,K) \cdot \boldsymbol{Z} \cdot ones(N,1) \\ &- \frac{1}{2} \sum_{k=1}^{K} \left[\operatorname{tr} \left((\boldsymbol{X} - \mu_{k} \cdot \boldsymbol{1}^{T})^{T} \boldsymbol{\Sigma}_{k}^{-1} (\boldsymbol{X} - \mu_{k} \cdot \boldsymbol{1}^{T}) \cdot \operatorname{diag}(\boldsymbol{z}_{k}) \right) \right] \end{split}$$

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

EM法

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

[M步骤]fix q(Z), change heta

fix
$$q \equiv q^k = p\left(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^k
ight) = p\left(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{
m old}
ight.
ight)$$

$$egin{aligned} heta^{k+1} &\equiv rgmax \mathcal{L}(q^k, oldsymbol{ heta}) \ &= rgmax \sum_{oldsymbol{ heta}} q^k \ln \left\{ rac{p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta})}{q^k}
ight\} \ &= rgmax \sum_{oldsymbol{Z}} q^k \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) - \sum_{oldsymbol{Z}} q^k \ln (q^k) \ &= rgmax \sum_{oldsymbol{Z}} q^k \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \ &\mathcal{L}(q^k, oldsymbol{ heta}^{k+1}) > \mathcal{L}(q^k, oldsymbol{ heta}^k) = \ln p(oldsymbol{X} | oldsymbol{ heta}^k) \end{aligned}$$

where $oldsymbol{ heta}$ old $\equiv oldsymbol{ heta}^k$

$$oldsymbol{Q}\left(oldsymbol{ heta},oldsymbol{ heta}^{
m old}
ight) \equiv \sum_{oldsymbol{Z}} q^k \ln p(oldsymbol{X},oldsymbol{Z}|oldsymbol{ heta})$$

[E步骤]fix θ , change q(Z)

fix $oldsymbol{ heta} \equiv oldsymbol{ heta}^{k+1}$

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

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

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

更新 θ old ,保留此时的 θ 值

$$oldsymbol{ heta}^{
m old} \ \Leftarrow oldsymbol{ heta}^{k+1}$$
 $q^{k+1} \equiv rgmax \mathcal{L}(q, oldsymbol{ heta}^{k+1})$ $\equiv rgmin \operatorname{KL}(q|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1})) = 0$ $= p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1})$ $\operatorname{ln} p(oldsymbol{X}|oldsymbol{ heta}^{k+1}) \equiv \mathcal{L}(q, oldsymbol{ heta}^{k+1}) + \operatorname{KL}(q|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1}))$ $= \mathcal{L}(q^k, oldsymbol{ heta}^{k+1}) + \operatorname{KL}(q^k|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1}))$ $= \mathcal{L}(q^{k+1}, oldsymbol{ heta}^{k+1})$ $\operatorname{ln} p(oldsymbol{X}|oldsymbol{ heta}^{k+1}) = \mathcal{L}(q^{k+1}, oldsymbol{ heta}^{k+1}) > \mathcal{L}(q^k, oldsymbol{ heta}^{k+1})$

in all

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

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

Calculation of [M 步骤]fix q(Z), change heta

fix
$$q \equiv q^k = p\left(oldsymbol{Z} | oldsymbol{X}, oldsymbol{ heta}^k
ight) = p\left(oldsymbol{Z} | oldsymbol{X}, oldsymbol{ heta}
ight.$$

where $oldsymbol{ heta}$ old $\equiv oldsymbol{ heta}^k$

$$oldsymbol{Q}\left(oldsymbol{ heta},oldsymbol{ heta}^{
m old}
ight) \equiv \sum_{oldsymbol{Z}} q^k \ln p(oldsymbol{X},oldsymbol{Z}|oldsymbol{ heta})$$

Here

$$egin{aligned} heta^{k+1} &\equiv rgmax \mathcal{L}(q^k, oldsymbol{ heta}) \ &= rgmax \sum_{oldsymbol{ heta}} q^k \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) - \sum_{oldsymbol{Z}} q^k \ln (q^k) \ &= rgmax \sum_{oldsymbol{ heta}} q^k \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \ &= rgmax oldsymbol{Q} \left(oldsymbol{ heta}, oldsymbol{ heta}
ight. \end{aligned}$$

Now find close form of Q

 $\sum_{m{Z}} p\left(m{Z}|m{X},m{ heta}^{
m old}
ight.)$ is [probability distribution] random variable **of** $m{Z}$ [probability distribution] random variable

$$\begin{split} \text{Here } q^k &= p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^k\right) = p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \\ \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) \left\{ (\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) \right. \\ &\left. - \frac{1}{2} \sum_{k=1}^K \left[\operatorname{tr}\left((\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) \cdot \operatorname{diag}(\boldsymbol{z}_k) \right) \right] \right\} \\ &= \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 \boldsymbol{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 - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_k) \boldsymbol{z}_{kn} \right] \right\} \\ &= \left[\boldsymbol{\gamma} \cdot \boldsymbol{1}. \right]^T (\ln \boldsymbol{\pi} - \frac{1}{2}|\boldsymbol{\Sigma}| - \frac{l}{2} \ln(2\pi) \cdot \boldsymbol{1}.) \\ &- \frac{1}{2} \sum_{k=1}^K \operatorname{tr} \left(\operatorname{diag}(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) \right) \end{split}$$

Here

$$p(oldsymbol{z_n}|oldsymbol{x_n},oldsymbol{ heta}^{ ext{old}}) = rac{p(oldsymbol{x_n}|oldsymbol{z_n},oldsymbol{ heta}^{ ext{old}})p(oldsymbol{z_n}|oldsymbol{ heta}^{ ext{old}})}{\sum\limits_{oldsymbol{z_n}}p(oldsymbol{x_n}|oldsymbol{z_n},oldsymbol{ heta}^{ ext{old}})p(oldsymbol{z_n}|oldsymbol{ heta}^{ ext{old}})}$$

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

$$egin{aligned} p(oldsymbol{z_n} = k | oldsymbol{x_n}, oldsymbol{ heta}^{ ext{old}}) &= rac{oldsymbol{\phi}_k^{ ext{old}}(oldsymbol{x_n}) \pi_k^{ ext{old}}}{\sum\limits_{k=1}^K oldsymbol{\phi}_k^{ ext{old}}(oldsymbol{x_n}) \pi_k^{ ext{old}}} &\equiv \gamma_{kn} \ oldsymbol{\gamma} &= oldsymbol{\phi}^{ ext{old}}(X) \odot (\pi_k^{ ext{old}} \cdot 1.^T) \odot (rac{1.}{1. \cdot (oldsymbol{\pi}^{ ext{old}})^T oldsymbol{\phi}^{ ext{old}}(X)}) \end{aligned}$$

maximize Q

$$egin{aligned} \max_{oldsymbol{ heta}} & oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta} ext{ old }
ight) \equiv \sum_{oldsymbol{Z}} p\left(oldsymbol{Z} | oldsymbol{X}, oldsymbol{ heta} ext{ old }
ight) \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \\ & = \left[oldsymbol{\gamma} \cdot 1.
ight]^T (\ln oldsymbol{\pi} - rac{1}{2} | oldsymbol{\Sigma} | - rac{l}{2} \ln (2\pi) \cdot 1.) \\ & - rac{1}{2} \sum_{k=1}^K \mathrm{tr} \Big(diag(oldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T \Sigma_k^{-1} (X - \mu_k \cdot 1^T) \Big) \\ & \mathrm{s.t.} \qquad 1.^T \cdot oldsymbol{\pi} = 1 \end{aligned}$$

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

for π

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

SO

$$[\gamma \cdot 1.] = \lambda \pi$$

Hence

$$egin{aligned} \lambda &= 1.^T ig[m{\gamma} \cdot 1. ig] = [1.^T m{\gamma} ig] \cdot 1. = ones(1, N) \cdot 1. = N \ m{\pi} &= rac{ig[m{\gamma} \cdot 1. ig]}{1.^T ig[m{\gamma} \cdot 1. ig]} = rac{ig[m{\gamma} \cdot 1. ig]}{N} \end{aligned}$$

for μ_k

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

SO

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

 $(X - \mu_k \cdot 1^T) \boldsymbol{\gamma}_k = 0$

Hence

$$\mu_k = rac{Xoldsymbol{\gamma}_k}{1.^Toldsymbol{\gamma}_k}$$

for Σ_k

$$\begin{split} dL &= \left[\boldsymbol{\gamma}_k^T \cdot 1 \cdot \right]^T (\frac{-1}{2}) \boldsymbol{\Sigma}_k^{-1} d\boldsymbol{\Sigma}_k \\ &- \frac{1}{2} \mathrm{tr} \Big(diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T d\boldsymbol{\Sigma}_k^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) \Big) \\ &= \left[\boldsymbol{\gamma}_k^T \cdot 1 \cdot \right]^T (\frac{-1}{2}) \boldsymbol{\Sigma}_k^{-1} d\boldsymbol{\Sigma}_k \\ &- \frac{1}{2} \mathrm{tr} \Big((\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T d\boldsymbol{\Sigma}_k^{-1} \Big) \\ &= -(\frac{1}{2}) \left[\boldsymbol{\gamma}_k^T \cdot 1 \cdot \right]^T \boldsymbol{\Sigma}_k^{-1} d\boldsymbol{\Sigma}_k \\ &+ \frac{1}{2} \mathrm{tr} \Big((\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \boldsymbol{\Sigma}_k^{-1} d\boldsymbol{\Sigma}_k \boldsymbol{\Sigma}_k^{-1} \Big) \\ &= -(\frac{1}{2}) \left[\boldsymbol{\gamma}_k^T \cdot 1 \cdot \right]^T \boldsymbol{\Sigma}_k^{-1} d\boldsymbol{\Sigma}_k \\ &+ \frac{1}{2} \mathrm{tr} \Big(\boldsymbol{\Sigma}_k^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \boldsymbol{\Sigma}_k^{-1} d\boldsymbol{\Sigma}_k \Big) \end{split}$$

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

$$\begin{split} \frac{\partial L}{\partial \Sigma_k} &= -(\frac{1}{2}) \big[\boldsymbol{\gamma}_k^T \cdot \boldsymbol{1}. \big] \boldsymbol{\Sigma}_k^{-T} \\ &+ \frac{1}{2} \Big(\boldsymbol{\Sigma}_k^{-T} (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \boldsymbol{\Sigma}_k^{-T} \Big) \\ &= 0 \end{split}$$

So

$$\begin{split} \left[\boldsymbol{\gamma}_k^T \cdot 1. \right] I_{l \times l} &= \boldsymbol{\Sigma}_k^{-T} (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \\ \left[\boldsymbol{\gamma}_k^T \cdot 1. \right] I_{l \times l} &= (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T \boldsymbol{\Sigma}_k^{-1} \\ \boldsymbol{\Sigma}_k &= \frac{(\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T) diag(\boldsymbol{\gamma}_k^T) \cdot (\boldsymbol{X} - \boldsymbol{\mu}_k \cdot \boldsymbol{1}^T)^T}{\left[\boldsymbol{\gamma}_k^T \cdot 1. \right]} \end{split}$$

To sum up

$$egin{aligned} oldsymbol{\pi} &= rac{\left[oldsymbol{\gamma} \cdot 1.
ight]}{1.^T \left[oldsymbol{\gamma} \cdot 1.
ight]} = rac{\left[oldsymbol{\gamma} \cdot 1.
ight]}{N} \ \mu_k &= rac{Xoldsymbol{\gamma}_k}{1.^T oldsymbol{\gamma}_k} \ \Sigma_k &= rac{(X - \mu_k \cdot 1^T) diag(oldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T}{1.^T oldsymbol{\gamma}_k} \end{aligned}$$

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

$$egin{aligned} p(oldsymbol{z_n} = k | oldsymbol{x_n}, oldsymbol{ heta}^{ ext{old}}) &= rac{oldsymbol{\phi}_k^{ ext{old}}(oldsymbol{x_n}) \pi_k^{ ext{old}}}{\sum\limits_{k=1}^K oldsymbol{\phi}_k^{ ext{old}}(oldsymbol{x_n}) \pi_k^{ ext{old}}} &\equiv \gamma_{kn} \ oldsymbol{\gamma} &= oldsymbol{\phi}^{ ext{old}}(X) \odot (\pi_k^{ ext{old}} \cdot 1.^T) \odot (rac{1.}{1. \cdot (oldsymbol{\pi}^{ ext{old}})^T oldsymbol{\phi}^{ ext{old}}(X)}) \end{aligned}$$

here γ_k is (N x 1) matrix

$$oldsymbol{\gamma} = \left[egin{array}{c} dots \ oldsymbol{\gamma}_k^T \ dots \end{array}
ight]$$