## Gaussian Mixture Model (GMM)

Here we set following symbols

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

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

 $\pi_k : 1 \times 1 \text{ (k = 1 - K)}$ 

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

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

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

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

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

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

 $Z:K \times N$ 

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

$$z_{kn}$$
: 1 × 1 (k = 1 ~ K)

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

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

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

Goal:

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

think about the Expectation-maximization algorithm

$$rg \max_{m{\pi}, m{\Sigma}, m{\mu}} \quad \ln p(m{X} | m{ heta}) = \mathcal{L}(q, m{ heta}) + \mathrm{KL}(q \| p)$$

where

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

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

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

$$egin{aligned} p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}, oldsymbol{\Sigma}, oldsymbol{\mu}) p(oldsymbol{Z} | oldsymbol{\pi}, oldsymbol{\Sigma}, oldsymbol{\mu}) p(oldsymbol{Z} | oldsymbol{\pi}) \ &= \prod_{n=1}^N p\left(oldsymbol{x_n} | oldsymbol{Z}, oldsymbol{\Sigma}, oldsymbol{\mu}
ight) \prod_{n=1}^N p\left(oldsymbol{z_n} | oldsymbol{\pi}
ight) \ &= \prod_{n=1}^N p\left(oldsymbol{x_n} | oldsymbol{z_n}, oldsymbol{\Sigma}, oldsymbol{\mu}
ight) \prod_{n=1}^N p\left(oldsymbol{z_n} | oldsymbol{\pi}
ight) \end{aligned}$$

Hence

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

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

EM method

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

[M step] fix q(Z), change  $\theta$ 

$$egin{aligned} ext{fix } q &\equiv q^k = p\left(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{ ext{old}}
ight) = p\left(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{ ext{old}}
ight) \ &= rgmax \sum_{oldsymbol{Z}} q^k \ln \left\{ rac{p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta})}{q^k} 
ight\} \ &= rgmax \sum_{oldsymbol{Z}} q^k \ln p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta}) - \sum_{oldsymbol{Z}} q^k \ln (q^k) \ &= rgmax \sum_{oldsymbol{Q}} q^k \ln p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta}) \ &= rgmax oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta}^{ ext{old}}
ight) \ &= rgmax oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta}^{ ext{old}}
ight) \ &= rgmax oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta}^{ ext{old}}
ight) \ &= \ln p(oldsymbol{X}|oldsymbol{ heta}^k) \end{aligned}$$

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

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

## [E step] fix $\theta$ , change q(Z)

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

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

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

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

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

update  $oldsymbol{ heta}$  old , to keep current  $oldsymbol{ heta}$  value

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

in all

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

So

$$egin{aligned} \lim_{k o \infty} \ln p(oldsymbol{X} | oldsymbol{ heta}^k) &= \lim_{k o \infty} \mathcal{L}(q^k, oldsymbol{ heta}^{k+1}) = \max_{oldsymbol{ heta}} \ln p(oldsymbol{X} | oldsymbol{ heta}) \ oldsymbol{ heta}_{max} &\equiv \max_{oldsymbol{ heta}} \ln p(oldsymbol{X} | oldsymbol{ heta}) \ &= \lim_{k o \infty} lpha \max_{oldsymbol{ heta}} oldsymbol{\sum}_{oldsymbol{Z}} q^k \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \ &= \lim_{k o \infty} rgmax oldsymbol{Q} \left(oldsymbol{ heta}, oldsymbol{ heta}^k 
ight) \end{aligned}$$

## Calculation of [M step] fix q(Z), change $\theta$

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

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

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

Here

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

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** Z [probability distribution] random variable

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

$$egin{aligned} oldsymbol{Q}\left(oldsymbol{ heta},oldsymbol{ heta}^{
m old}
ight) &\equiv \sum_{oldsymbol{Z}} p\left(oldsymbol{Z}|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \ln p(oldsymbol{X},oldsymbol{Z}|oldsymbol{ heta}) &= \sum_{oldsymbol{Z}} p\left(oldsymbol{Z}|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \left\{ \left(\ln oldsymbol{\pi}^T - rac{1}{2}|oldsymbol{\Sigma}|^T
ight) \cdot oldsymbol{Z} \cdot ones(N,1) \\ &- rac{1}{2} \sum_{k=1}^K \left[ \mathrm{tr} \Big( (X - \mu_k \cdot \mathbf{1}^T)^T \Sigma_k^{-1} (X - \mu_k \cdot \mathbf{1}^T) \cdot \mathrm{diag}(oldsymbol{z}_k) \Big) 
ight] \Big\} \\ &= \left\{ \left[ \sum_{n=1}^N \sum_{oldsymbol{Z}_n} p\left(oldsymbol{z}_n|oldsymbol{x}_n, oldsymbol{ heta}^{
m old}
ight) oldsymbol{z}_n^T \Big] (\ln oldsymbol{\pi} - rac{1}{2}|oldsymbol{\Sigma}| - rac{1}{2} \ln(2\pi) \cdot 1.) 
ight\} \\ &= \left[ oldsymbol{\gamma} \cdot \mathbf{1}. 
ight]^T (\ln oldsymbol{\pi} - rac{1}{2}|oldsymbol{\Sigma}| - rac{1}{2} \ln(2\pi) \cdot 1.) \\ &- rac{1}{2} \sum_{k=1}^K \mathrm{tr} \Big( diag(oldsymbol{\gamma}_k^T) \cdot (X - \mu_k \cdot \mathbf{1}^T)^T \Sigma_k^{-1} (X - \mu_k \cdot \mathbf{1}^T) \Big) \end{aligned}$$

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

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

maximize Q

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

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

for  $\pi$ 

$$rac{\partial L}{\partial oldsymbol{\pi}} = ig[oldsymbol{\gamma} \cdot 1.ig] \odot rac{1.}{oldsymbol{\pi}} - \lambda 1. = 0$$

SO

$$egin{bmatrix} m{\gamma} \cdot 1. \end{bmatrix} = \lambda m{\pi}$$

Hence

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

for  $\mu_k$ 

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

SO

$$egin{aligned} (X - \mu_k \cdot 1^T) diag(oldsymbol{\gamma}_k^T) \cdot 1. &= 0 \ (X - \mu_k \cdot 1^T) oldsymbol{\gamma}_k &= 0 \end{aligned}$$

Hence

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

for  $\Sigma_k$ 

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

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

$$egin{aligned} rac{\partial L}{\partial \Sigma_k} &= -(rac{1}{2})ig[m{\gamma}_k^T \cdot 1.ig] \Sigma_k^{-T} \ &+ rac{1}{2} \Big(\Sigma_k^{-T} (X - \mu_k \cdot 1^T) diag(m{\gamma}_k^T) \cdot (X - \mu_k \cdot 1^T)^T \Sigma_k^{-T}\Big) \ &= 0 \end{aligned}$$

So

$$egin{aligned} egin{aligned} egi$$

## To sum up

For M step

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

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

For E step

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

here  $\gamma_k$  is (N x 1) matrix

$$oldsymbol{\gamma} = egin{bmatrix} drstrucking \ oldsymbol{\gamma}_k^T \ drstrucking \ \end{pmatrix}$$