EM Formula Notes for PRKM(Probabilistic Reduced K-means)

- Equal Variance in mixture model -

Model :

$$p(\mathbf{x}) = \sum_{k=1}^{K} p_k(z) p_k(\mathbf{x}\,|\,z)$$

Where,

- Z is hidden variable

$Z = \{0,1\}$

- $\mathbf{x} = \mathbf{A}\mathbf{y}_k + \mathbf{v}$
- $\mathbf{A}$ is $p \times q$ matrix.
- $\mathbf{y} \sim N(\mu_k, \sigma^2 \mathbf{I})$
- $\mathbf{v} \sim N(0, \mathbf{R})$, $\mathbf{R} = \epsilon \mathbf{I}$

$$P_k(\mathbf{x}\,|\,Z) = \int_y p_k(\mathbf{x}\,|\,Z, \mathbf{y}) p(\mathbf{y}) d\mathbf{y} = N(\mathbf{A}\mu_k, \mathbf{A}(\sigma^2 I)\mathbf{A}^\top + \epsilon \mathbf{I})$$

& $p_k(\mathbf{x}\,|\,\mathbf{y}_k, Z) = N(\mathbf{A}\mathbf{y}_k, \epsilon \mathbf{I})$

My original model include σ^2 parameter(variance in reduced space).

However, σ^2 suffered identifiability problem, that is σ^2 is estimated too small.

Perhaps, most of the error(variance of reduced space) is captured by ϵ in practical computation. So, from now on, consider $\sigma^2 = 1$.

EM Derive

$$ln\left(p(\mathbf{X})\right) = ln\left(\sum_Z p(\mathbf{X}, \mathbf{Z})\right) = ln\left(\sum_Z q(\mathbf{Z}) \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})}\right)$$

$$\geq \sum_{Z} q(\mathbf{Z}) ln \left(\frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} \right) = \sum_{Z} q(\mathbf{Z}) ln \left(p(\mathbf{X}, \mathbf{Z}) \right) - q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

$$= \sum_{Z} q(\mathbf{Z}) ln \left(\int_{\mathbf{Y}} p(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) d\mathbf{Y} \right) - q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

$$= \sum_{Z} q(\mathbf{Z}) ln \left(\int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) \frac{p(\mathbf{X}, \mathbf{Y}, \mathbf{Z})}{q(\mathbf{Y} | \mathbf{Z})} d\mathbf{Y} \right) - q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

$$\geq \sum_{Z} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(\frac{p(\mathbf{X}, \mathbf{Y}, \mathbf{Z})}{q(\mathbf{Y} | \mathbf{Z})} \right) d\mathbf{Y} - q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

$$= \sum_{Z} q(\mathbf{Z}) \left[\int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(p(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \right) - q(\mathbf{Y} | \mathbf{Z}) ln \left(q(\mathbf{Y} | \mathbf{Z}) \right) d\mathbf{Y} \right] - \sum_{\mathbf{Z}} q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

$$= \sum_{Z} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(p(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \right) d\mathbf{Y} - \sum_{\mathbf{Z}} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(q(\mathbf{Y} | \mathbf{Z}) \right) d\mathbf{Y} - \sum_{\mathbf{Z}} q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

$$= E_{\mathbf{Y}, \mathbf{Z}} \left[ln \left(p(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \right) \right] - \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(q(\mathbf{Y} | \mathbf{Z}) \right) d\mathbf{Y} - \sum_{\mathbf{Z}} q(\mathbf{Z}) ln \left(q(\mathbf{Z}) \right)$$

The 2nd and 3rd term is entropy, and only the first term contains $\mathbf{X}$. So we

$$ln \left(p(\mathbf{X}) \right) = ln \sum_{Z} \int_{\mathbf{Y}} p(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) d\mathbf{Y}$$

$$= ln \sum_{\mathbf{Z}} \int_{\mathbf{Y}} \frac{q(\mathbf{Y} | \mathbf{Z}) p(\mathbf{Y} | \mathbf{X}, \mathbf{Z}) p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Y} | \mathbf{Z})} d\mathbf{Y} = ln \sum_{\mathbf{Z}} \int_{\mathbf{Y}} \frac{q(\mathbf{Z}) q(\mathbf{Y} | \mathbf{Z}) p(\mathbf{Y} | \mathbf{X}, \mathbf{Z}) p(\mathbf{Z} | \mathbf{X}) p(\mathbf{X})}{q(\mathbf{Z}) q(\mathbf{Y} | \mathbf{Z})} d\mathbf{Y}$$

$$\geq \sum_{\mathbf{Z}} q(\mathbf{Z}) ln \int_{\mathbf{Y}} \frac{q(\mathbf{Y} | \mathbf{Z}) p(\mathbf{Y} | \mathbf{X}, \mathbf{Z}) p(\mathbf{Z} | \mathbf{X}) p(\mathbf{X})}{q(\mathbf{Z}) q(\mathbf{Y} | \mathbf{Z})} d\mathbf{Y}$$

$$\geq \sum_{\mathbf{Z}} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(\frac{p(\mathbf{Y} | \mathbf{X}, \mathbf{Z}) p(\mathbf{Z} | \mathbf{X}) p(\mathbf{X})}{q(\mathbf{Z}) q(\mathbf{Y} | \mathbf{Z})} d\mathbf{Y} \right)$$

$$= \sum_{\mathbf{Z}} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) \left[ln \left(\frac{p(\mathbf{Z} | \mathbf{X})}{q(\mathbf{Z})} \right) + ln \left(\frac{p(\mathbf{Y} | \mathbf{X}, \mathbf{Z})}{q(\mathbf{Y} | \mathbf{Z})} \right) + ln \left(p(\mathbf{X}) \right) \right] d\mathbf{Y}$$

$$= \sum_{\mathbf{Z}} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(p(\mathbf{X}) \right) d\mathbf{Y} + \sum_{\mathbf{Z}} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(\frac{p(\mathbf{Z} | \mathbf{X})}{q(\mathbf{Z})} \right) d\mathbf{Y} + \sum_{\mathbf{Z}} q(\mathbf{Z}) \int_{\mathbf{Y}} q(\mathbf{Y} | \mathbf{Z}) ln \left(\frac{p(\mathbf{Y} | \mathbf{X}, \mathbf{Z})}{q(\mathbf{Y} | \mathbf{Z})} \right) d\mathbf{Y}$$

$$= ln\left(p(\mathbf{X})\right)\sum_{\mathbf{Z}} q(\mathbf{Z})\int_{\mathbf{Y}} q(\mathbf{Y}|\mathbf{Z})d\mathbf{Y} - \sum_{\mathbf{Z}} q(\mathbf{Z})ln\left(\frac{q(\mathbf{Z})}{p(\mathbf{Z}|\mathbf{X})}\right) - \sum_{\mathbf{Z}} q(\mathbf{Z})\int_{\mathbf{Y}} q(\mathbf{Y}|\mathbf{Z})ln\left(\frac{q(\mathbf{Y}|\mathbf{Z})}{p(\mathbf{Y}|\mathbf{X},\mathbf{Z})}\right)d\mathbf{Y}$$

$$= ln\left(p(\mathbf{X})\right) - \sum_{\mathbf{Z}} q(\mathbf{Z})ln\left(\frac{q(\mathbf{Z})}{p(\mathbf{Z}|\mathbf{X})}\right) - \sum_{\mathbf{Z}} q(\mathbf{Z})\int_{\mathbf{Y}} q(\mathbf{Y}|\mathbf{Z})ln\left(\frac{q(\mathbf{Y}|\mathbf{Z})}{p(\mathbf{Y}|\mathbf{X},\mathbf{Z})}\right)d\mathbf{Y}$$

Thus, if we set $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X})$ and $q(\mathbf{Y}|\mathbf{Z}) = p(\mathbf{Y}|\mathbf{X},\mathbf{Z})$, the log-likelihood lower bound becomes the same as the log-likelihood, and that means for every EM iteration we can guarantee increasing log-likelihood.

EM Algorithm Formular

$$\theta = \{\pi_1, .., \pi_K \mu_1, .. \mu_K, \sigma, \mathbf{A}, \epsilon\}$$
$$l(\theta) = ln\prod_{i=1}^{N}\sum_{k=1}^{K} p_k(\mathbf{x}_i, \mathbf{y}_i, z_i) = \sum_{i=1}^{N} ln\sum_{k=1}^{K} p_k(z_i)p_k(\mathbf{x}_i, \mathbf{y}_i|z_i) = \sum_{i=1}^{N} ln\sum_{k=1}^{K} p_k(z_i)p_k(\mathbf{y}_i|z_i)p_k(\mathbf{x}_i|\mathbf{y}_i, z_i)$$
$$= \sum_{i=1}^{N} ln\sum_{k=1}^{K}\pi_k\frac{1}{(2\pi)^{q/2}|\sigma^2\mathbf{I}_q|^{1/2}}exp\left(-\frac{1}{2\sigma_k^2}\|\mathbf{y}_{ik}-\mu_k\|^2\right)\frac{1}{(2\pi)^{p/2}|\epsilon\mathbf{I}_p|^{1/2}}exp\left(-\frac{1}{2\epsilon}\|\mathbf{x}_i - \mathbf{A}\mathbf{y}_{ik}\|^2\right)$$

E-step

We fill in the missing value by their expectations.

1. Fill in Z

We fill $\mathbf{Z}$ using the $p_k(\mathbf{Z}|\mathbf{X})$

$$r_k(\mathbf{x}_i) = p(z_{ik} = 1|\mathbf{x}_i) = \frac{\hat{\pi}_k p(\mathbf{x}_i|z_{ik} = 1)}{\sum_{j=1}^{K}\hat{\pi}_j p(\mathbf{x}_i|z_{ij} = 1)}$$

2. Fill in Y

We fill $\mathbf{Y}_1, ..., \mathbf{Y}_K$ by using $E\left(p(\mathbf{Y}|\mathbf{X})\right)$

$$\begin{pmatrix}\mathbf{X}|\mathbf{Z}\\ \mathbf{Y}|\mathbf{Z}\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{A}\mu_\mathbf{k}\\ \mu_\mathbf{k}\end{pmatrix}, \begin{pmatrix}\mathbf{A}(\sigma^2\mathbf{I})\mathbf{A}^\top + \epsilon\mathbf{I} & \sigma_k^2\mathbf{A}\\ \sigma^2\mathbf{A}^\top & \sigma^2\mathbf{I}\end{pmatrix}\right)$$

Here,
$$Cov(\mathbf{Y}|\mathbf{Z}, \mathbf{X}|\mathbf{Z}) = E(\mathbf{Y}\mathbf{X}^\top|\mathbf{Z}) - E(\mathbf{Y}|\mathbf{Z})E(\mathbf{X}^\top|\mathbf{Z})$$
$$= E(\mathbf{Y}(\mathbf{A}\mathbf{Y})^\top|\mathbf{Z}) - \mu_k(\mathbf{A}\mu_\mathbf{k})^\top = E(\mathbf{Y}\mathbf{Y}^\top\mathbf{A}^\top|\mathbf{Z}) - \mu_k(\mathbf{A}\mu_\mathbf{k})^\top = E(\mathbf{Y}\mathbf{Y}^\top|\mathbf{Z})\mathbf{A}^\top - \mu_k\mu_k^\top\mathbf{A}^\top$$
$$= \left(Var(\mathbf{Y}|\mathbf{Z}) + E(\mathbf{Y}|\mathbf{Z})E(\mathbf{Y}^\top|\mathbf{Z})\right)\mathbf{A}^\top - \mu_k\mu_k^\top\mathbf{A}^\top = \sigma^2\mathbf{I}\mathbf{A}^\top + \mu_k\mu_k^\top\mathbf{A}^\top - \mu_k\mu_k^\top\mathbf{A}^\top$$
$$= \sigma^2\mathbf{I}\mathbf{A}^\top$$

Then,

$$\mathbf{y}\,|\,\mathbf{x},z \sim N\left(\mu_k + \sigma^2\mathbf{A}^\top\mathbf{M}^{-1}(\mathbf{x}-\mathbf{A}\mu_k), \sigma^2\mathbf{I} - \sigma^2\mathbf{A}^\top\mathbf{M}^{-1}\sigma^2\mathbf{A}\right),$$

$$E(\mathbf{y}\,|\,\mathbf{x},z) = \mu_k + \sigma^2\mathbf{A}^\top\mathbf{M}^{-1}(\mathbf{x}-\mathbf{A}\mu_k) \ = \mathbf{e}_{ki}$$

$$E(\mathbf{y}\mathbf{y}^\top\,|\,\mathbf{x},z) = Var(\mathbf{y}\,|\,\mathbf{x},z) + E(\mathbf{y}\,|\,\mathbf{x},z)E(\mathbf{y}^\top\,|\,\mathbf{x},z) = \left(\sigma^2\mathbf{I} - \sigma^2\mathbf{A}^\top\mathbf{M}^{-1}\sigma^2\mathbf{A}\right) + E(\mathbf{y}\,|\,\mathbf{x},z)E(\mathbf{y}^\top\,|\,\mathbf{x},z) = \mathbf{V}_{ki}$$

Here, $\mathbf{M} = \mathbf{A}(\sigma^2\mathbf{I})\mathbf{A}^\top + \epsilon\mathbf{I}$

By using this value we take expectation of $l(\theta)$ with respect to expected values

$$E_{\mathbf{Y},\mathbf{Z}}(l(\theta)) = E\left[\sum_{i=1}^{N} ln \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{q/2}\,|\,\sigma^2\mathbf{I}_q\,|^{1/2}} exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}_{ik}-\mu_k\|^2\right) \frac{1}{(2\pi)^{p/2}\,|\,\epsilon\mathbf{I}_p\,|^{1/2}} exp\left(-\frac{1}{2\epsilon}\|\mathbf{x}_i - \mathbf{A}\mathbf{y}_{ik}\|^2\right)\right]$$

$$= \sum_{i=1}^{N} ln \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{q/2}\,|\,\sigma^2\mathbf{I}_q\,|^{1/2}} exp\left(-\frac{1}{2\sigma^2}E\left[(\mathbf{y}_{ik}-\mu_k)^\top(\mathbf{y}_{ik}-\mu_k)\right]\right) \frac{1}{(2\pi)^{p/2}\,|\,\epsilon\mathbf{I}_p\,|^{1/2}} exp\left(-\frac{1}{2\epsilon}E\left[(\mathbf{x}_i - \mathbf{A}\mathbf{y}_{ik})^\top(\mathbf{x}_i - \mathbf{A}\mathbf{y}_{ik})\right]\right)$$

Here,

$$E\left[\mathbf{y}_{ik}^\top\mathbf{y}_{ik} - 2\mathbf{y}_{ik}^\top\mu_k + \mu_k^\top\mu_k\right] = E\left[\mathbf{y}_{ik}^\top\mathbf{y}_{ik}\right] - 2\left[\mathbf{y}_{ik}\right]^\top\mu_k + \mu_k^\top\mu_k$$

$$E\left[\mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}\mathbf{y}_{ik} + \mathbf{y}_{ik}^\top\mathbf{A}^\top\mathbf{A}\mathbf{y}_{ik}\right] = \mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}E\left[\mathbf{y}_{ik}\right] + E\left[tr(\mathbf{A}\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\mathbf{A}^\top)\right]$$
$$= \mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top\mathbf{A})$$

Then,

$$E_{\mathbf{Y},\mathbf{Z}}(l(\theta)) = \sum_{i=1}^{N} ln \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{q/2}\,|\,\sigma^2\mathbf{I}_q\,|^{1/2}} exp\left[-\frac{1}{2\sigma^2}\left(E\left[tr(\mathbf{y}_{ik}\mathbf{y}_{ik}^\top)\right]) - 2\left[\mathbf{y}_{ik}\right]^\top\mu_k + \mu_k^\top\mu_k\right)\right] \frac{1}{(2\pi)^{p/2}\,|\,\epsilon\mathbf{I}_p\,|^{1/2}} exp\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top\mathbf{A}))\right)\right]$$

M-step

1. Update $\pi_j\,(j = 1,...,K)$

$$\frac{\partial}{\partial\pi_j}\left(l(\theta) + \lambda(\sum_{k=1}^{K}\pi_k - 1)\right) = \sum_{i=1}^{N}\frac{const_j}{\sum_{k=1}^{K}\pi_k \times const_k} + \lambda := 0$$

$\rightarrow \displaystyle\sum_{k=1}^{K}\sum_{i=1}^{N}\frac{\pi_j \times const_j}{\sum_{k=1}^{K}\pi_k \times const_k} + \pi_j\lambda := 0$, and here $\lambda = -N$

$$\sum_{i=1}^{N} r_j(\mathbf{x}_i) = N\pi_j \rightarrow \pi_j = \frac{\sum_{i=1}^{N} r_j(\mathbf{x}_i)}{N}$$

2. Update $\mu_j\,(j = 1,...,K)$

$$\frac{\partial E(l(\theta))}{\partial \mu_j} = \frac{\partial}{\partial \mu_j}\sum_{i=1}^{N} ln \sum_{k=1}^{K} const_k \times exp\left(-\frac{1}{2\sigma^2}(tr(\mathbf{V}_{ki}) - 2\mathbf{e}_{ki}^\top \mu_k + \mu_k^\top \mu_k\right)$$

$$= \sum_{i=1}^{N} \frac{const_j \times exp\left(-\frac{1}{2\sigma^2}(tr(\mathbf{V}_{ki}) - 2\mathbf{e}_{ki}^\top \mu_k + \mu_k^\top \mu_k\right)}{\sum_{k=1}^{K} const_k \times exp\left(-\frac{1}{2\sigma^2}(tr(\mathbf{V}_{ki}) - 2\mathbf{e}_{ki}^\top \mu_k + \mu_k^\top \mu_k\right)} \frac{\partial}{\partial \mu_j}\left(-\frac{1}{2\sigma^2} tr(\mathbf{V}_{ji}) - 2\mathbf{e}_{ji}^\top \mu_j + \mu_j^\top \mu_j\right)$$

$$= \sum_{i=1}^{N} r_j(\mathbf{x}_i)\left(\frac{1}{\sigma^2}\mathbf{e}_{ji} - \frac{1}{\sigma^2}\mu_j\right) := 0, \text{ then}$$

$$\sum_{i=1}^{N} r_j(\mathbf{x}_i)\frac{1}{\sigma^2}\mathbf{e}_{ji} = \sum_{i=1}^{N} r_j(\mathbf{x}_i)\frac{1}{\sigma^2}\mu_j \rightarrow \mu_j = \frac{\sum_{i=1}^{N} r_j(\mathbf{x}_i)\mathbf{e}_{ji}}{\sum_{i=1}^{N} r_j(\mathbf{x}_i)}$$

3. Update $\mathbf{A}$

$$\frac{\partial}{\partial \mathbf{A}}\sum_{i=1}^{N} ln \sum_{k=1}^{K} const_k \times exp\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A})\right)\right]$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} \frac{const_k \times exp\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A}))\right)\right] \frac{\partial}{\partial \mathbf{A}}\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A}))\right)\right]}{\sum_{k=1}^{K} const_k \times exp\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A})\right)\right]}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} r_k(\mathbf{x}_i)\frac{\partial}{\partial \mathbf{A}}\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A}))\right)\right] = \sum_{i=1}^{N}\sum_{k=1}^{K} r_k(\mathbf{x}_i)\left[-\frac{1}{2\epsilon}\left(-2\mathbf{x}_i\mathbf{e}_{ki}^\top + 2\mathbf{A}\mathbf{V}_{ki}\right)\right] := 0$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} r_k(\mathbf{x}_i)\left(\mathbf{A}\mathbf{V}_{ki}\right) = \sum_{i=1}^{N}\sum_{k=1}^{K} r_k(\mathbf{x}_i)\left(\mathbf{x}_i\mathbf{e}_{ki}^\top\right)$$

$$\mathbf{A} = \sum_{i=1}^{N}\sum_{k=1}^{K}\left(r_k(\mathbf{x}_i)\mathbf{x}_i\mathbf{e}_{ki}^\top\right)\left(\sum_{i=1}^{N}\sum_{k=1}^{K} r_k(\mathbf{x}_i)\mathbf{V}_{ki}\right)^{-1}$$

4. Update ϵ

$$\frac{\partial}{\partial \epsilon}\sum_{i=1}^{N} ln \sum_{k=1}^{K} const_k \times \frac{1}{(2\pi)^{p/2}|\epsilon \mathbf{I}_p|^{1/2}} exp\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A})\right)\right]$$

let $a = \frac{1}{(2\pi)^{p/2}|\epsilon \mathbf{I}_p|^{1/2}}$ & $b = exp\left[-\frac{1}{2\epsilon}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A})\right)\right]$

$\frac{\partial a}{\partial \epsilon} = \frac{1}{(2\pi)^{p/2}|\epsilon \mathbf{I}_p|^{1/2}}\left(-\frac{p}{2\epsilon}\right)$ &

$$\frac{\partial b}{\partial \epsilon} = b \times \left(\frac{1}{2\epsilon^2}\left(\mathbf{x}_i^\top \mathbf{x}_i - 2\mathbf{x}_i^\top \mathbf{A} E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top \mathbf{A})\right)\right)$$

Thus,

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{const_k \times ab\left(-\frac{p}{2\epsilon}\right) + const_k \times ab\left(\frac{1}{2\epsilon^2}\left(\mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top\mathbf{A})\right)\right)}{\sum_{k=1}^{K} const_k \times ab}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} r(\mathbf{x}_i)\left(-\frac{p}{2\epsilon}\right) + r(\mathbf{x}_i)\left(\frac{1}{2\epsilon^2}\left(\mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}E\left[\mathbf{y}_{ik}\right] + tr(E\left[\mathbf{y}_{ik}\mathbf{y}_{ik}^\top\right]\mathbf{A}^\top\mathbf{A})\right)\right) := 0$$

$$\epsilon = \frac{\sum_{i=1}^{N}\sum_{k=1}^{K} r(\mathbf{x}_i)\left(\mathbf{x}_i^\top\mathbf{x}_i - 2\mathbf{x}_i^\top\mathbf{A}\mathbf{e}_{ik} + tr(\mathbf{V}_{ik}\mathbf{A}^\top\mathbf{A})\right)}{Np}$$