EM algorithm

1 Algorithm

1.1 Preparation

Put $S_k[1] = \sum_{n=1}^{N} \gamma_{nk}, \quad S_k[\boldsymbol{x}] = \sum_{n=1}^{N} \gamma_{nk}\boldsymbol{x_n}, \quad S_k\left[\boldsymbol{x}\boldsymbol{x}^T\right] = \sum_{n=1}^{N} \gamma_{nk}\boldsymbol{x}_n\boldsymbol{x}_n^T.$

I will derive the update formulas of the parameters $\boldsymbol{\pi}$,$\boldsymbol{\mu}$,$\boldsymbol{\Lambda}$(p.22) by letting the partial derivative of the lower bound(p.20) w.r.t. each parameter equal to zero.

M step maximizes $\log p(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})$ to maximaize lower bound $\int q(\boldsymbol{Z}) \log p(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta}) dZ$.

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X}}[\log p(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X}}[\ln \left\{ \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{nk}} \right\}] \\
&= \mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X}}[\sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \left(\ln \pi_k + \ln N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)\right)] \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \mathrm{E}[z_{nk}] \left(\ln \pi_k + \ln N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)\right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left(\ln \pi_k + \ln N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)\right) \\
&= F
\end{aligned}
$$

The extreme values of F under the condition of $\sum_{k=1}^{K} \pi_k = 1$ is derived by Lagrange multipliers method.

$$
F' = F + \lambda(\sum_{k=1}^{K} \pi_k - 1)
$$

is defined.

1.1.1 $\boldsymbol{\pi}$

$$
\begin{aligned}
\frac{\partial}{\partial \pi_k} F' &= \sum_{n=1}^{N} \gamma_{nk} \frac{\partial}{\partial \pi_k} \ln \pi_k + \lambda \\
&= \sum_{n=1}^{N} \gamma_{nk} \frac{1}{\pi_k} + \lambda \qquad (1) \\
&= 0 \qquad (2)
\end{aligned}
$$

By (1)(2),

$$
\begin{aligned}
\lambda \pi_k &= - \sum_{n=1}^{N} \gamma_{nk} \qquad (3) \\
\lambda \sum_{k=1}^{K} \pi_k &= - \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma_{nk} \\
\lambda &= - \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma_{nk} \qquad (4)
\end{aligned}
$$

Substitute (4) into (3),

$$
\begin{aligned}
\pi_k &= \frac{\sum_{n=1}^{N} \gamma_{nk}}{\sum_{k=1}^{K} \sum_{n=1}^{N} \gamma_{nk}} \\
&= \frac{S_k[1]}{S.[1]}
\end{aligned}
$$

1.1.2 $\boldsymbol{\mu}$

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\mu}_k} F' &= \sum_{n=1}^{N} \gamma_{nk} \frac{\partial}{\partial \boldsymbol{\mu}_k} \ln N \left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1} \right) \\
&= \sum_{n=1}^{N} \gamma_{nk} \frac{\partial}{\partial \boldsymbol{\mu}_k} \left[-\frac{1}{2} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\boldsymbol{x}_n - \boldsymbol{\mu}_k) \right] \\
&= \sum_{n=1}^{N} \gamma_{nk} \left\{ \boldsymbol{\Lambda}_k \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right) \right\} (\because \frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^T A \boldsymbol{x} = \left(A + A^T \right) \boldsymbol{x}) \\
&= \boldsymbol{\Lambda}_k \left\{ \sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n - \boldsymbol{\mu}_k \sum_{n=1}^{N} \gamma_{nk} \right\} \qquad (5) \\
&= 0 \qquad (6)
\end{aligned}
$$

By (5)(6),

$$\begin{aligned}\boldsymbol{\mu}_k &= \frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n}{\sum_{n=1}^{N} \gamma_{nk}} \\ &= \frac{S_k[\boldsymbol{x}]}{S_k[1]}\end{aligned} \tag{7}$$

1.1.3 $\boldsymbol{\Lambda}$

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\Lambda}_k} F' &= \sum_{n=1}^{N} \gamma_{nk} \frac{\partial}{\partial \boldsymbol{\Lambda}_k} \ln N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right) \\ &= \sum_{n=1}^{N} \gamma_{nk} \left[-\frac{1}{2}\boldsymbol{\Lambda}_k + \frac{1}{2}\left(\boldsymbol{\Lambda}_k \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Lambda}_k\right)\right] && (8) \\ &= 0 && (9)\end{aligned}$$

By (8)(9),

$$\begin{aligned}\boldsymbol{\Lambda}_k \sum_{n=1}^{N} \gamma_{nk} &= \sum_{n=1}^{N} \gamma_{nk} \left(\boldsymbol{\Lambda}_k \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Lambda}_k\right) \\ \boldsymbol{\Lambda}_k \sum_{n=1}^{N} \gamma_{nk} &= \boldsymbol{\Lambda}_k \left(\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n \boldsymbol{x}_n^T - \sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n \boldsymbol{\mu}_k^T - \boldsymbol{\mu}_k \sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n^T - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \sum_{n=1}^{N} \gamma_{nk}\right) \boldsymbol{\Lambda}_k \\ \boldsymbol{\Lambda}_k &= \boldsymbol{\Lambda}_k \left(\frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n \boldsymbol{x}_n^T}{\sum_{n=1}^{N} \gamma_{nk}} - \frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n}{\sum_{n=1}^{N} \gamma_{nk}} \boldsymbol{\mu}_k^T - \boldsymbol{\mu}_k \frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n^T}{\sum_{n=1}^{N} \gamma_{nk}} + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T\right) \boldsymbol{\Lambda}_k \\ \boldsymbol{\Lambda}_k &= \boldsymbol{\Lambda}_k \left(\frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n \boldsymbol{x}_n^T}{\sum_{n=1}^{N} \gamma_{nk}} - \boldsymbol{\mu}_k {\boldsymbol{\mu}_k}^T - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T + \boldsymbol{\mu}_k {\boldsymbol{\mu}_k}^T\right) \boldsymbol{\Lambda}_k \quad (\because (7)) \\ \boldsymbol{\Lambda}_k^{-1} &= \frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n \boldsymbol{x}_n^T}{\sum_{n=1}^{N} \gamma_{nk}} - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \\ \boldsymbol{\Lambda}_k^{-1} &= \frac{S_k[\boldsymbol{x}\boldsymbol{x}^T]}{S_k[1]} - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T\end{aligned}$$

1.2 EM algorithm

I will explain outline of EM algorithm. EM algorithm is a deterministic algorithm for ML estimation. Suppose $\boldsymbol{X}$: observed variables and $\boldsymbol{Z}$: latent variables, we aim to estimate the parameters $\boldsymbol{\pi}$,$\boldsymbol{\mu}$,$\boldsymbol{\Lambda}$ of Finite Gaussian Mixture Models.

E-step calculates a posterior distribution over latent variables $\boldsymbol{Z}$ by the following equation;

$$\begin{aligned}q^*\left(z_{nk} = 1\right) &= \frac{\pi_k N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)}{\sum_{k'=1}^{K} \pi_{k'} N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_{k'}, \boldsymbol{\Lambda}_{k'}^{-1}\right)} \\ &= \gamma_{nk}\end{aligned}$$

M-step updates parameters $\boldsymbol{\pi}$,$\boldsymbol{\mu}$,$\boldsymbol{\Lambda}$. Ratio $\boldsymbol{\pi}^*$ is calculated by the the following equation;

$$\pi_k^* = \frac{S_k[1]}{S.[1]}$$

Mean $\boldsymbol{\mu}^*$ is calculated by the following equation;

$$\boldsymbol{\mu}_k^* = \frac{S_k[\boldsymbol{x}]}{S_k[1]}$$

Variance Λ^* is calculated by the following equation;

$$\boldsymbol{\Lambda}_k^{-1^*} = \frac{S_k[\boldsymbol{x}\boldsymbol{x}^T]}{S_k[1]} - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^T$$

.

E-step and M-step are iterated until convergence of probability distribution below threshold.

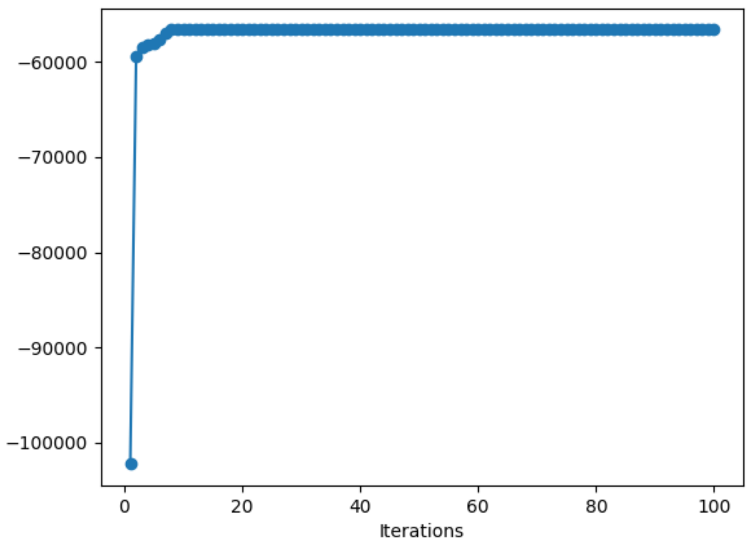
2 Result

2.1 The value of the log likelihood

The log likelihood function is

$$\log p(\boldsymbol{X};\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Lambda}) = \sum_{n=1}^{N} \log\left\{\sum_{k=1}^{K} \pi_k N\left(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k\right)\right\}$$

The value of the log likelihood is



2.2 Classification

I classify each point into the class with the highest value of γ_{nk}. This result is shown below.