VB algorithm

1 Algorithm

1.1 Preparation

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

Derive the variational posteriors of the parameters $\boldsymbol{\pi}$,$\boldsymbol{\mu}$,$\boldsymbol{\Lambda}$ (p.47) by using the formulas (p.46)

$$
\begin{aligned}
\log q^*(\boldsymbol{\pi}) &= \log p(\boldsymbol{\pi}) + \langle \log p(\boldsymbol{Z}|\boldsymbol{\pi}) \rangle_{q(\boldsymbol{Z})} + \text{ const.} \\
&= \log \left(\frac{\Gamma\left(\sum_{k=1}^{K} \alpha_{0k}\right)}{\prod_{k=1}^{K} \Gamma\left(\alpha_{0k}\right)} \prod_{k=1}^{K} \pi_k^{\alpha_{0k}-1} \right) + \langle \log \left(\prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \right) \rangle_{q(\boldsymbol{Z})} + \text{ const.} \\
&= \sum_{k=1}^{K} (\alpha_{0k} - 1) \log \pi_k + \sum_{n=1}^{N} \sum_{k=1}^{K} \langle z_{nk} \rangle_{q(\boldsymbol{Z})} \log \pi_k + \text{ const.} \\
&= \sum_{k=1}^{K} (\alpha_{0k} - 1) \log \pi_k + \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \log \pi_k + \text{ const.} \\
&= \sum_{k=1}^{K} \left(\sum_{n=1}^{N} \gamma_{nk} \log \pi_k + \alpha_{0k} - 1 \right) \log \pi_k + \text{ const.} \\
&= \log \left(\prod_{k=1}^{K} \pi_k^{\left(\sum_{n=1}^{N} \gamma_{nk}\right) + \alpha_{0k} - 1} \right) + \text{ const}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
q^*(\boldsymbol{\pi}) &= \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}) \\
\alpha_k &= \alpha_{0k} + \sum_{n=1}^{N} \gamma_{nk} \\
&= \alpha_{0k} + S_k[1]
\end{aligned}
$$

$$\begin{aligned}
\log q^{*}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \log p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) + \langle \log p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \rangle_{q(\boldsymbol{Z})} + \text{ const.} \\
&= \log \left(\prod_{k=1}^{K} N\left(\boldsymbol{\mu}_k | \boldsymbol{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}\right) W\left(\boldsymbol{\Lambda}_k | \mathbf{W}_0, v_0\right) \right) \\
&+ \langle \log \left(\prod_{n=1}^{N} \prod_{k=1}^{K} N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{nk}} \right) \rangle_{q(\boldsymbol{Z})} + \text{ const.} \\
&= \sum_{k=1}^{K} \log N\left(\boldsymbol{\mu}_k | \boldsymbol{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}\right) W\left(\boldsymbol{\Lambda}_k | \mathbf{W}_0, v_0\right) \\
&+ \sum_{k=1}^{K} \langle \sum_{n=1}^{N} \log N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{nk}} \rangle_{q(\boldsymbol{Z})} + \text{ const.}
\end{aligned}$$

$$\begin{aligned}
\log q^{*}(\boldsymbol{\mu_k}, \boldsymbol{\Lambda_k}) &= \log N\left(\boldsymbol{\mu}_k | \boldsymbol{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}\right) W\left(\boldsymbol{\Lambda}_k | \mathbf{W}_0, v_0\right) \\
&+ \langle \sum_{n=1}^{N} \log N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{nk}} \rangle_{q(\boldsymbol{Z})} + \text{ const.} \\
&= \log N\left(\boldsymbol{\mu}_k | \boldsymbol{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}\right) + \log W\left(\boldsymbol{\Lambda}_k | \mathbf{W}_0, v_0\right) \\
&+ \sum_{n=1}^{N} \gamma_{nk} \log N\left(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right) + \text{ const.} \\
&= -\frac{\beta_0}{2} \left(\boldsymbol{\mu}_k - \boldsymbol{m}_0\right)^T \boldsymbol{\Lambda}_k \left(\boldsymbol{\mu}_k - \boldsymbol{m}_0\right) + \frac{1}{2} \ln |\boldsymbol{\Lambda}_k| - \frac{1}{2} \operatorname{Tr}\left(\boldsymbol{\Lambda}_k \mathbf{W}_0^{-1}\right) \\
&+ \frac{(\nu_0 - D - 1)}{2} \ln |\boldsymbol{\Lambda}_k| - \frac{1}{2} \sum_{n=1}^{N} \gamma_{nk} \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Lambda}_k \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right) \\
&+ \frac{1}{2} \left(\sum_{n=1}^{N} \gamma_{nk} \right) \ln |\boldsymbol{\Lambda}_k| + \text{ const.}
\end{aligned} \tag{1}$$

By (1),

$$\log q^{\star}\left(\boldsymbol{\mu}_k | \boldsymbol{\Lambda}_k\right) = -\frac{1}{2} \boldsymbol{\mu}_k^T \left[\beta_0 + \sum_{n=1}^{N} \gamma_{nk} \right] \boldsymbol{\Lambda}_k \boldsymbol{\mu}_k + \boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k \left[\beta_0 \boldsymbol{m}_0 + \sum_{n=1}^{N} \gamma_{n_k} \boldsymbol{x}_n \right] + \text{ const.}$$

$$q^{\star}\left(\boldsymbol{\mu}_k | \boldsymbol{\Lambda}_k\right) = N\left(\boldsymbol{\mu}_k | \boldsymbol{m}_k, (\beta_k \boldsymbol{\Lambda}_k)^{-1}\right)$$

where

$$\begin{aligned}
\beta_k &= \beta_0 + \sum_{n=1}^{N} \gamma_{nk} \\
&= \beta_0 + S_k[1] \\
\boldsymbol{m}_k &= \frac{1}{\beta_k}\left(\beta_0 \boldsymbol{m}_0 + \sum_{n=1}^{N} \gamma_{nk}\boldsymbol{x}_n\right) \\
&= \frac{1}{\beta_k}\left(\beta_0 \boldsymbol{m}_0 + S_k[\boldsymbol{x}]\right) \\
&= \frac{\beta_0 \boldsymbol{m}_0 + S_k[\boldsymbol{x}]}{\beta_0 + S_k[1]}
\end{aligned}$$

By $\log q^\star(\boldsymbol{\Lambda}_k) = \log q^*(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) - \log q^\star(\boldsymbol{\mu}_k|\boldsymbol{\Lambda}_k)$,

$$\begin{aligned}
\log q^\star(\boldsymbol{\Lambda}_k) = & -\frac{\beta_0}{2}(\boldsymbol{\mu}_k - \boldsymbol{m}_0)^T \boldsymbol{\Lambda}_k (\boldsymbol{\mu}_k - \boldsymbol{m}_0) + \frac{1}{2}\log|\boldsymbol{\Lambda}_k| - \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Lambda}_k \mathbf{W}_0^{-1}\right) \\
& + \frac{(\nu_0 - D - 1)}{2}\log|\boldsymbol{\Lambda}_k| - \frac{1}{2}\sum_{n=1}^{N}\gamma_{nk}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\boldsymbol{x}_n - \boldsymbol{\mu}_k) \\
& - \frac{1}{2}\ln|\boldsymbol{\Lambda}_k| + \text{ const.} \\
& = \frac{(\nu_k - D - 1)}{2}\log|\boldsymbol{\Lambda}_k| - \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Lambda}_k \mathbf{W}_k^{-1}\right) + \text{ const.} \\
& q^\star(\boldsymbol{\Lambda}_k) = W(\boldsymbol{\Lambda}_k|\mathbf{W}_k, \nu_k)
\end{aligned}$$

Here,

$$\begin{aligned}
\mathbf{W}_k^{-1} &= \mathbf{W}_0^{-1} + \beta_0(\boldsymbol{\mu}_k - \boldsymbol{m}_0)(\boldsymbol{\mu}_k - \boldsymbol{m}_0)^T + \sum_{n=1}^{N}\gamma_{nk}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T \\
&- \beta_k(\boldsymbol{\mu}_k - \boldsymbol{m}_k)(\boldsymbol{\mu}_k - \boldsymbol{m}_k)^T \\
&= \boldsymbol{W}_0^{-1} + \beta_0 \boldsymbol{m}_0 \boldsymbol{m}_0^T + S_k\left[\boldsymbol{x}\boldsymbol{x}^T\right] - \beta_k \boldsymbol{m}_k \boldsymbol{m}_k^T \\
\nu_k &= \nu_0 + \sum_{n=1}^{N}\gamma_{nk} \\
&= \nu_0 + S_k[1]
\end{aligned}$$

1.2 VB algorithm

I will explain outline of VB algorithm. VB algorithm is a bayesian extension of the EM algorithm. 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 the variational posterior over latent variables $\boldsymbol{Z}$ by the following equation;

$$\gamma_{nk} \propto \widetilde{\pi}_k \widetilde{\Lambda}_k^{1/2} \exp\left\{-\frac{D}{2\beta_k} - \frac{\nu_k}{2}(\boldsymbol{x}_n - \boldsymbol{m}_k)^T \mathbf{W}_k (\boldsymbol{x}_n - \boldsymbol{m}_k)\right\}$$

where

$$\log \widetilde{\Lambda}_k = \sum_{i=1}^{D} \psi \left(\frac{\nu_k + 1 - i}{2} \right) + D \log 2 + \log |\mathbf{W}_k|$$
$$\log \tilde{\pi}_k = \psi \left(\alpha_k \right) - \psi(\widehat{\alpha})$$

M-step updates parameters by the the following equation;

$$\alpha_k = \alpha_{0k} + \sum_{n=1}^{N} \gamma_{nk}$$
$$\beta_k = \beta_0 + S_k[1]$$
$$\boldsymbol{m}_k = \frac{\beta_0 \boldsymbol{m}_0 + S_k[\boldsymbol{x}]}{\beta_0 + S_k[1]}$$
$$\mathbf{W}_k^{-1} = \boldsymbol{W}_0^{-1} + \beta_0 \boldsymbol{m}_0 \boldsymbol{m}_0^T + S_k \left[\boldsymbol{x}\boldsymbol{x}^T \right] - \beta_k \boldsymbol{m}_k \boldsymbol{m}_k^T$$
$$\nu_k = \nu_0 + S_k[1]$$

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

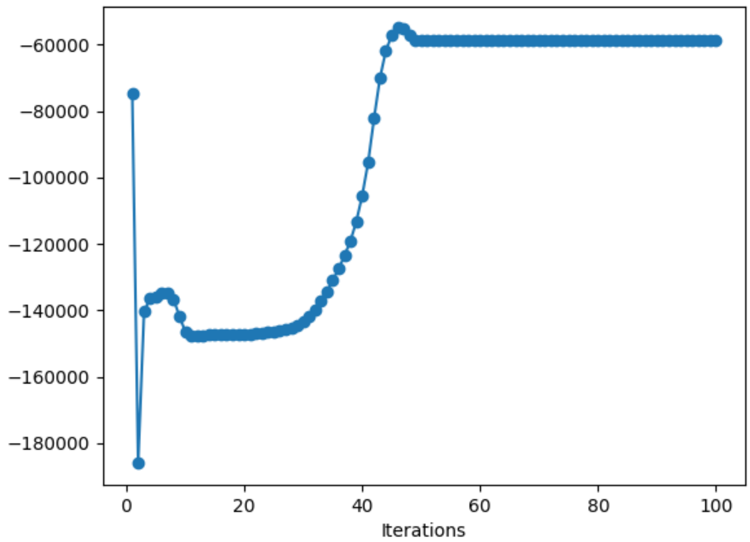
2 Result

2.1 The value of the log likelihood

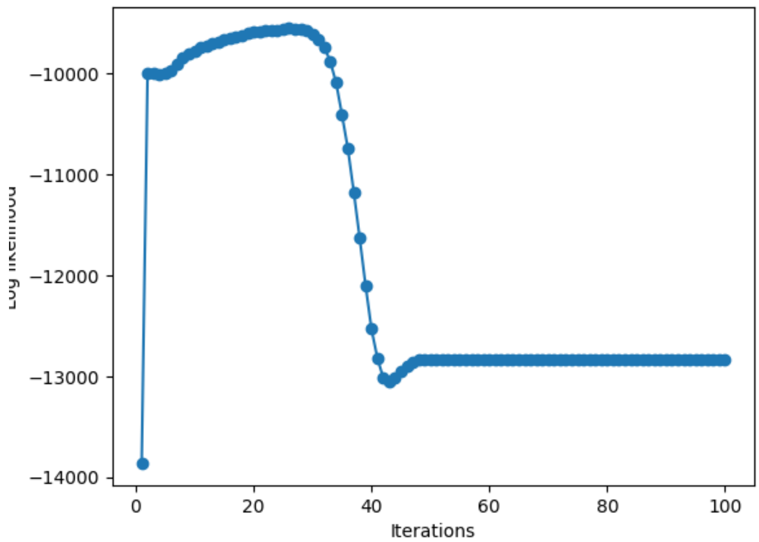
The loglikelihood functions are

$$\log p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log N \left(x_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1} \right) \tag{2}$$
$$\log p(\boldsymbol{Z}|\boldsymbol{\pi}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log \pi_k \tag{3}$$

The value of the log likelihood(2) is



The value of the log likelihood(3) is



2.2 Classification

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