Sample of GMM parameters estimation using Gibbs sampling method

Ryo Ozaki
Ritsumeikan University
Graduate School of Information Science and Engineering
Emergent Systems Laboratory
ryo.ozaki@em.ci.ritsumei.ac.jp

April 23, 2018

1 Graphical model

This section shows the graphical model of Gaussian mixture model (GMM) which were used this paper.

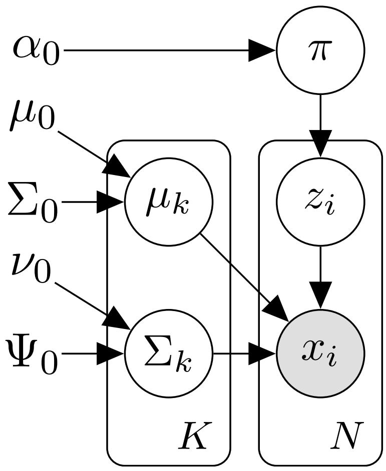

Figure 1: Graphical model of GMM.

The probabilistic generative process of GMM is shown below.

$$
\begin{aligned}
\lambda &= \{\ \alpha_0, \mu_0, \Sigma_0, \nu_0, \Psi_0\ \} && (1)\\
\theta &= \{\ \pi, z_{1:N}, \mu_{1:K}, \Sigma_{1:K}\ \} && (2)
\end{aligned}
$$

$$
\begin{aligned}
\pi &\sim \mathrm{Dir}(\pi|\alpha_0) && (3)\\
\mu_k &\sim \mathrm{N}(\mu_k|\mu_0, \Sigma_0) && (4)\\
\Sigma_k &\sim \mathrm{W}^{-1}(\Sigma_k|\nu_0, \Psi_0) && (5)\\
z_i &\sim \mathrm{Cat}(z_i|\pi) && (6)\\
x_i &\sim \mathrm{N}(x_i|\mu_{z_i}, \Sigma_{z_i}) && (7)
\end{aligned}
$$

$$
\begin{aligned}
k &= 1, 2, \ldots, K && (8)\\
i &= 1, 2, \ldots, N && (9)
\end{aligned}
$$

Here, λ represents the set of hyperparameters, θ represents the set of parameters, Dir represents the Dirichlet distribution, N represents the normal distribution, W^{-1} represents the inverse-Wishart distribution, Cat represents the categorical distribution.

2 Posterior distribution

This section shows posterior distributions of the θ.

2.1 Posterior distribution of z_i

The posterior distribution of z_i is represented as follows.

$$
\begin{aligned}
p(z_i = k|x_i, \pi, \mu_{1:K}, \Sigma_{1:K}) &\propto p(x_i|z_i = k, \pi, \mu_{1:K}, \Sigma_{1:K}) \\
&\qquad \times p(z_i = k|\pi, \mu_{1:K}, \Sigma_{1:K}) && (10)\\
&\propto p(x_i|z_i = k, \mu_k, \Sigma_k)p(z_i = k|\pi) && (11)\\
&= \mathrm{N}(x_i|\mu_k, \Sigma_k)\mathrm{Cat}(\mathrm{z_i = k}|\pi) && (12)\\
&= \mathrm{Cat}(z_i = k|\hat{\pi}) && (13)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\pi}_k^- &= \mathrm{N}(x_i|\mu_k, \Sigma_k) \times \pi_k && (14)\\
\hat{\pi}_k &= \frac{\hat{\pi}_k^-}{\sum_{l=1}^{K} \hat{\pi}_l^-} && (15)
\end{aligned}
$$

2.2 Posterior distribution of π

The posterior distribution of π is represented as follows.

$$
\begin{aligned}
p(\pi|\alpha_0, z_{1:N}) &\propto p(z_{1:N}|\pi, \alpha_0)p(\pi|\alpha_0) && (16)\\
&\propto p(z_{1:N}|\pi)p(\pi|\alpha_0) && (17)\\
&= \prod_{i=1}^{N}\{p(z_i|\pi)\}p(\pi|\alpha_0) && (18)\\
&= \prod_{i=1}^{N}\{\mathrm{Cat}(z_i|\pi)\}\mathrm{Dir}(\pi|\alpha_0) && (19)\\
&= \mathrm{Mult}(m|\pi)\mathrm{Dir}(\pi|\alpha_0) && (20)\\
&= \mathrm{Dir}(\pi|\hat{\alpha}) && (21)
\end{aligned}
$$

$$
\begin{aligned}
m_k &= \sum_{i=1}^{N}\delta(z_i = k) && (22)\\
\delta(\mathrm{CONDITION}) &= \begin{cases} 0 & (\mathrm{CONDITION\ is\ false}) \\ 1 & (\mathrm{CONDITION\ is\ true}) \end{cases} && (23)\\
\hat{\alpha} &= m + \alpha_0 && (24)
\end{aligned}
$$

Here, Mult represents multinomial distribution.

2.3 Posterior distribution of Σ_k

The posterior distribution of Σ_k is represented as follows.

$$
\begin{aligned}
p(\Sigma_k|\nu_0, \Psi_0, x_{1:N}, z_{1:N}, \mu_k) &= p(\Sigma_k|\nu_0, \Psi_0, x^k_{1:M}, \mu_k) && (25)\\
&\propto p(x^k_{1:M}|\nu_0, \Psi_0, \mu_k, \Sigma_k)p(\Sigma_k|\nu_0, \Psi_0, \mu_k) && (26)\\
&\propto p(x^k_{1:M}|\mu_k, \Sigma_k)p(\Sigma_k|\nu_0, \Psi_0) && (27)\\
&= \prod_{j=1}^{M}\left\{p(x^k_j|\mu_k, \Sigma_k)\right\}p(\Sigma_k|\nu_0, \Psi_0) && (28)\\
&= \prod_{j=1}^{M}\left\{\mathrm{N}(x^k_j|\mu_k, \Sigma_k)\right\}\mathrm{W}^{-1}(\Sigma_k|\nu_0, \Psi_0) && (29)\\
&= \mathrm{W}^{-1}(\Sigma_k|\hat{\nu}, \hat{\Psi}) && (30)
\end{aligned}
$$

$$
\begin{aligned}
x^k_{1:M} &= \{\ x_i \mid z_i = k, i = 1, \ldots, N\ \} && (31)\\
m_k &= \sum_{i=1}^{N}\delta(z_i = k) && (32)\\
\hat{\nu} &= m_k + \nu_0 && (33)\\
\hat{\Psi} &= \sum_{j=1}^{M}\left\{(x_j - \mu_k)\cdot(x_j - \mu_k)^T\right\} + \Psi_0 && (34)
\end{aligned}
$$

2.4 Posterior distribution of μ_k

The posterior distribution of μ_k is represented as follows.

$$\begin{aligned}
p(\mu_k|\mu_0,\Sigma_0,x_{1:N},z_{1:N},\Sigma_k) &= p(\mu_k|\mu_0,\Sigma_0,x^k_{1:M},\Sigma_k) && (35)\\
&\propto p(x^k_{1:M}|\mu_0,\Sigma_0,\mu_k,\Sigma_k)p(\mu_k|\mu_0,\Sigma_0,\Sigma_k) && (36)\\
&\propto p(x^k_{1:M}|\mu_k,\Sigma_k)p(\mu_k|\mu_0,\Sigma_0) && (37)\\
x^k_{1:M} &= \{x_i|z_i=k, i=1,\ldots,N\} && (38)\\
&= \prod_{j=1}^{M}\left\{p(x^k_j|\mu_k,\Sigma_k)\right\}p(\mu_k|\mu_0,\Sigma_0) && (39)\\
&= \prod_{j=1}^{M}\left\{\mathrm{N}(x^k_j|\mu_k,\Sigma_k)\right\}\mathrm{N}(\mu_k|\mu_0,\Sigma_0) && (40)\\
&= \mathrm{N}(\mu_k|\hat{\mu},\hat{\Sigma}) && (41)
\end{aligned}$$

$$\begin{aligned}
x^k_{1:M} &= \{\ x_i \mid z_i=k, i=1,\ldots,N\ \} && (42)\\
m_k &= \sum_{i=1}^{N}\delta(z_i=k) && (43)\\
\hat{\Sigma} &= \left(m_k\Sigma_k^{-1}+\Sigma_0^{-1}\right)^{-1} && (44)\\
\hat{\mu} &= \hat{\Sigma}\cdot\left\{\Sigma_k^{-1}\cdot\sum_{j=1}^{M}\left(x^k_j\right)+\Sigma_0^{-1}\cdot\mu_0\right\} && (45)
\end{aligned}$$

3 Experiments

This section shows results of experiments.

3.1 Truth parameters

Parameters which were used in this experiments represented follows.

$$\begin{aligned}
\pi^* &= \begin{bmatrix} 0.7 & 0.2 & 0.1 \end{bmatrix} && (46)\\
\mu^*_{1:3} &= \begin{bmatrix} 3.0 \\ 0.0 \end{bmatrix}, \begin{bmatrix} -3.0 \\ 0.0 \end{bmatrix}, \begin{bmatrix} 0.0 \\ 5.0 \end{bmatrix} && (47)\\
\Sigma^*_{1:3} &= \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}, \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}, \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix} && (48)
\end{aligned}$$

3.2 Hyperparameters

Hyperparameters which were used in this experiments represented follows.

$$\alpha_0 = \begin{bmatrix} 1.0 & 1.0 & 1.0 \end{bmatrix} \quad (49)$$

$$\mu_0 = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} \quad (50)$$

$$\Sigma_0 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix} \quad (51)$$

$$\nu_0 = 2.0 \quad (52)$$

$$\Psi_0 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix} \quad (53)$$

3.3 Experiment 1

This section shows a clustering result of the dataset which has 300 observations. Figure 2 shows the dataset which are used in this experiment. In addition, we set $K = 4$ which is a number of clusters. We did Gibbs sampling 150 steps.

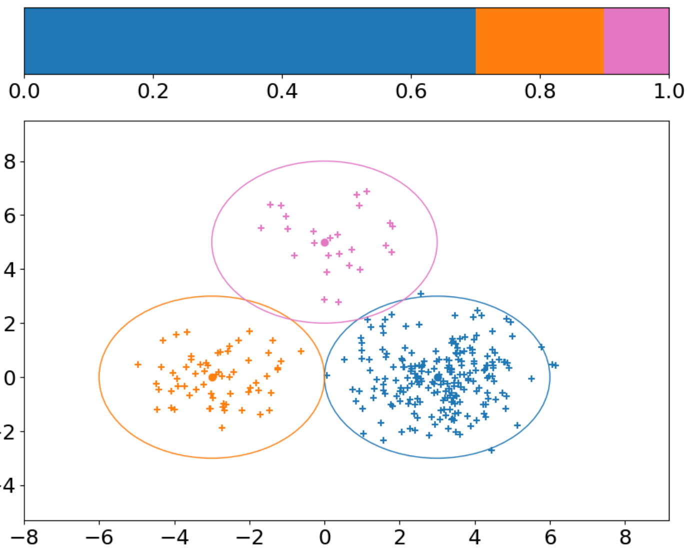

Figure 2: Truth data (300 points)

Figure 3 shows a result at 150-th step in Gibbs sampling. In addition, you can see the process of Gibbs sampling at the file named "GMM_Gibbs_result1.gif" which in this directory. Also, you can see the stochastic behavior of Gibbs sampling.

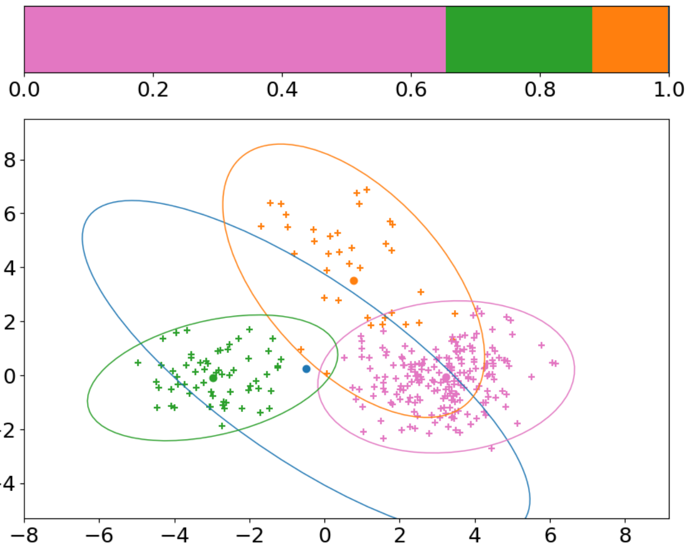

Figure 3: Result at 150-th step in Gibbs sampling

3.4 Experiment 2

This section shows a clustering result of the dataset which has 3000 observations. Figure 4 shows the dataset which are used in this experiment. Other conditions are same as is experiment 1.

Figure 5 shows a result at 150-th step in Gibbs sampling. In addition, you can see the process of Gibbs sampling at the file named "GMM_Gibbs_result2.gif" which in this directory.

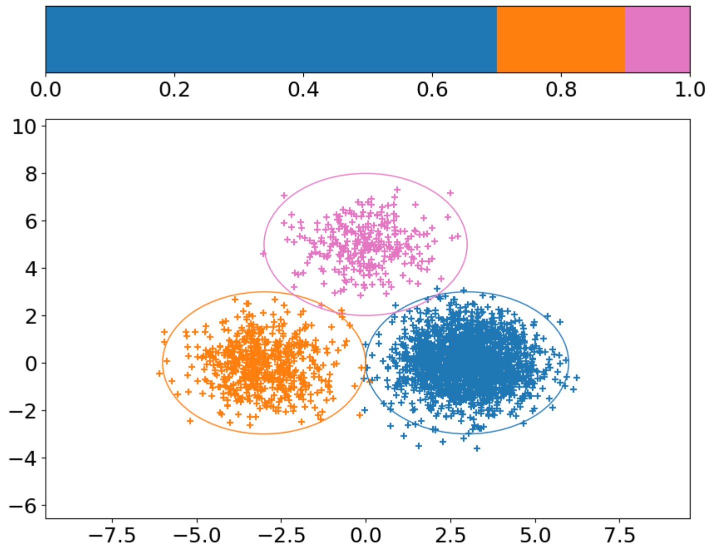

Figure 4: Truth data (3000 points)

4 Summary

This paper derived the posterior distributions of parameters θ. In addition, we estimated the parameters of synthesis dataset which has 300 and 3000 observations. Also, we showed the stochastic behavior of Gibbs sampling into "GMM_Gibbs_result1.gif" and "GMM_Gibbs_result2.gif."

You can the same experiment on your computer using this GitHub repository. If you want to do, please run following commands.

1. `git clone https://github.com/RyoOzaki/GibbsSampling`
   Clone this repository to your computer.

2. `cd GibbsSampling/GMM`
   Change directory to "GMM" in "GibbsSampling".

3. `python GMM_Gibbs.py`
   Run Gibbs sampling. You can get result in "tmp" directory.

4. `sh make_gif.sh`
   Convert images in "tmp" directory to "result.gif."

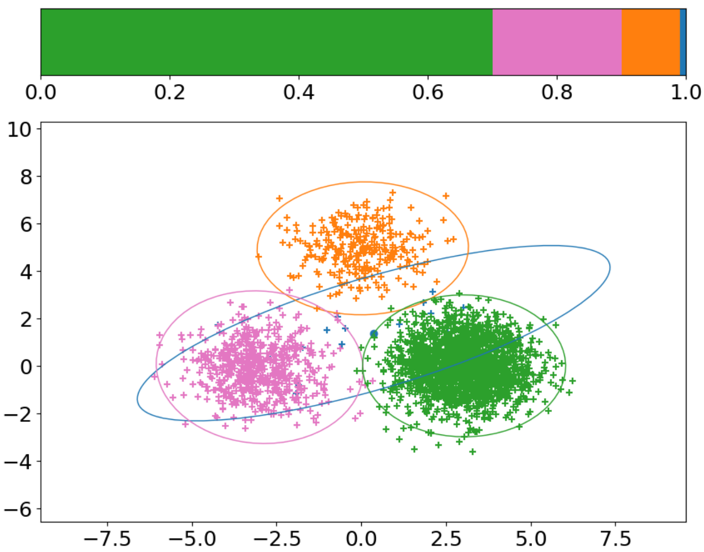

Figure 5: Result at 150-th step in Gibbs sampling