Gaussian Mixture

2017 年 5 月 14 日

要約

ガウス混合分布を Gibbs Sampling と Collapsed Gibbs Sampling で考える。

1 Data Generating Process

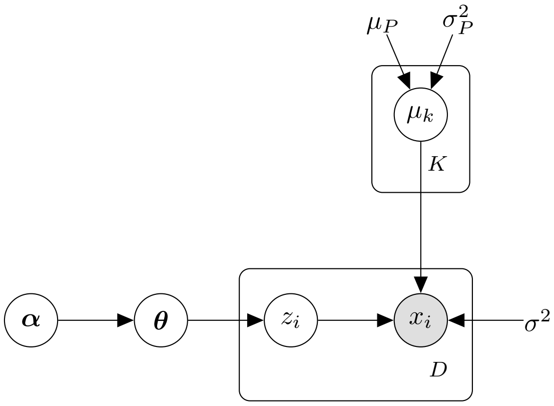

図 1 Gaussian Mixture

$$\mu_k \sim \mathcal{N}(\mu_{\mathrm{P}}, \sigma_{\mathrm{P}}^2), \quad k = 1, \cdots, K \tag{1}$$
$$\boldsymbol{\theta} \sim \mathrm{Dir}(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K) \tag{2}$$
$$z_i \sim \mathrm{Multi}(1|\boldsymbol{\theta}), \quad i = 1, \cdots, D \tag{3}$$
$$x_i \sim \mathcal{N}(u_{z_i=k}, \sigma^2), \quad i = 1, \cdots, D \tag{4}$$

2 Gibbs Sampling

2.1 全確率変数の結合分布

$$p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\mu}|\boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) \tag{5}$$
$$= p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}, \sigma^2) p(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\mu}|\boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) \tag{6}$$
$$= p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}, \sigma^2) p(\mathbf{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\alpha}) p(\boldsymbol{\mu}|\mu_P, \sigma_P^2) \tag{7}$$
$$= \prod_{i=1}^{D} \left\{ p(x_i|\mu_{z_i}, \sigma^2) p(z_i|\boldsymbol{\theta}) \right\} \cdot p(\boldsymbol{\theta}|\boldsymbol{\alpha}) \cdot \prod_{k=1}^{K} p(\mu_k|\mu_P, \sigma_P^2) \tag{8}$$

2.2 更新式

2.2.1 z_i のサンプリング

$$
\begin{aligned}
&p(z_i = k | x_i, \mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (9)\\
&\propto p(z_i = k, x_i, \mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (10)\\
&= p(x_i | z_i = k, \mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) \cdot p(z_i = k | \mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (11)\\
&= p(x_i | z_i = k, \mu_k, \sigma^2) \cdot p(z_i = k | \mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu}) \cdot p(\mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (12)\\
&= p(x_i | \mu_{z_i=k}, \sigma^2) \cdot p(z_i = k | \boldsymbol{\theta}) \cdot p(\mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \boldsymbol{\theta}, \boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (13)\\
&\propto \underbrace{p(x_i | \mu_{z_i=k}, \sigma^2)}_{\text{Normal Distribution}} \cdot \underbrace{p(z_i = k | \boldsymbol{\theta})}_{\theta_k} && (14)
\end{aligned}
$$

2.2.2 θ のサンプリング

$$
\begin{aligned}
&p(\boldsymbol{\theta} | \mathbf{x}, \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (15)\\
&\propto p(\boldsymbol{\theta}, \mathbf{x}, \mathbf{z}, \boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (16)\\
&= p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (17)\\
&= p(\mathbf{x} | \mathbf{z}, \boldsymbol{\mu}, \sigma^2) p(\mathbf{z} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \boldsymbol{\mu}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(\boldsymbol{\mu} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (18)\\
&\propto \underbrace{p(\mathbf{z} | \boldsymbol{\theta})}_{\text{Multi}} \underbrace{p(\boldsymbol{\theta} | \boldsymbol{\alpha})}_{\text{Dir}} && (19)\\
&\propto \prod_{k=i}^{K} \theta_k^{n_k + \alpha_k - 1}, \quad \text{where } n_k = \sum_{i=1}^{D} \delta(z_i = k) && (20)
\end{aligned}
$$

$\boldsymbol{\theta}$ は Dirichlet 分布からなので正規化項も考えると、

$$
\frac{\Gamma\left(\sum_{k=1}^{K} n_k + \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(n_k + \alpha_k)} \prod_{k=1}^{K} \theta_k^{n_k + \alpha_k - 1} \tag{21}
$$

2.2.3 μ_k のサンプリング

$$
\begin{aligned}
&p(\mu_k | \boldsymbol{\mu}^{\setminus k}, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (22)\\
&\propto p(\mu_k, \boldsymbol{\mu}^{\setminus k}, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (23)\\
&= p(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\mu}^{\setminus k}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(\boldsymbol{\mu}, \boldsymbol{\mu}^{\setminus k} \mathbf{z}, \boldsymbol{\theta} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (24)\\
&= p(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\mu}^{\setminus k}, \mathbf{z}, \sigma^2) p(\mu_k | \mu_P, \sigma_P^2) p(\boldsymbol{\mu}^{\setminus k}, \mathbf{z}, \boldsymbol{\theta} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (25)\\
&\propto \underbrace{p(\mathbf{x} | \boldsymbol{\mu}, \mathbf{z}, \sigma^2)}_{\text{Normal}} \underbrace{p(\mu_k | \mu_P, \sigma_P^2)}_{\text{Normal}} && (26)\\
&\propto \prod_{k=1}^{K} \left[\left\{ \prod_{i=1}^{D} \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma^2} \right) \cdot \delta(z_i = k) \right\} \times \left\{ \frac{1}{\sqrt{\sigma_P^2}} \exp\left(-\frac{(\mu_k - \mu_P)^2}{2\sigma_P^2} \right) \right\} \right] && (27)
\end{aligned}
$$

Wikipedia の Conjugate prior の記事を参考にすると、更新後の平均は、

$$
\left(\frac{\mu_P}{\sigma_P^2} + \frac{\sum_{i=1}^{D} x_i \cdot \delta(z_i = k)}{\sigma^2} \right) \Bigg/ \left(\frac{1}{\sigma_P^2} + \frac{\sum_{i=1}^{D} \delta(z_i = k)}{\sigma^2} \right), \quad k = 1, \cdots, K \tag{28}
$$

分散は、

$$\left(\frac{1}{\sigma_P^2} + \frac{\sum_{i=1}^{D} \delta(z_i = k)}{\sigma^2}\right)^{-1}, \quad k = 1, \cdots, K \tag{29}$$

となる。

3 Collapsed Gibbs Sampling

$\boldsymbol{\theta}$ と $\boldsymbol{\mu}$ を積分消去する。このときのグラフィカルモデルは以下のようになる。

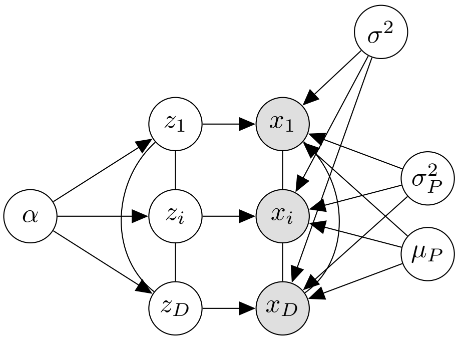

図 2 Collapsed Gibbs Sampling

周辺化されることで、隠れ変数やデータに依存関係が生まれていることに注意する必要がある。依存関係があるため、z を全て同時にサンプリングすることはできないが、$z_1, \cdots, z_D$ のように 1 つずつサンプリングすることは可能である。ある z_i は、それ以外のデータ $\mathbf{z}^{\backslash i}$ が与えられた状態でサンプリングする。

$$\begin{aligned}
&p(z_i = k | x_i, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (30)\\
&\propto p(z_i = k, x_i, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (31)\\
&= p(x_i | z_i = k, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(z_i = k, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (32)\\
&= p(x_i | z_i = k, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(z_i = k | \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(\mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i} | \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (33)\\
&\propto p(x_i | z_i = k, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) p(z_i = k | \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}, \mu_P, \sigma_P^2, \sigma^2) && (34)\\
&= p(x_i | z_i = k, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \mu_P, \sigma_P^2, \sigma^2) p(z_i = k | \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}) && (35)\\
&= \int p(x_i | z_i = k, \boldsymbol{\mu}, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \mu_P, \sigma_P^2, \sigma^2) p(\boldsymbol{\mu} | z_i = k, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \mu_P, \sigma_P^2, \sigma^2) d\boldsymbol{\mu} && (36)\\
&\qquad \times \int p(z_i = k | \boldsymbol{\theta}, \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}) p(\boldsymbol{\theta} | \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}) d\boldsymbol{\theta} && (37)\\
&= \int p(x_i | z_i = k, \mu_k, \sigma^2) p(\mu_k | \mathbf{x}^{\backslash i}, \mathbf{z}^{\backslash i}, \mu_P, \sigma_P^2, \sigma^2) d\mu_k \int p(z_i = k | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{z}^{\backslash i}, \boldsymbol{\alpha}) d\boldsymbol{\theta} && (38)\\
&= \mathcal{N}(x_i | \mu_{\text{New}}, \sigma^2) \cdot \frac{n_k^{\backslash i} + \alpha_k}{\sum_{k=1}^{K} n_k^{\backslash i} + \alpha_k} && (39)
\end{aligned}$$

式 (38) は、それぞれ x_i、z_i に関する予測分布 (posterior predictive distribution)[*1]であると解釈することができる。第 1 項は Normal と Normal、第 2 項は Multinomial (Categorical) と Dirichlet である。

[*1] Wikipedia, Coonjugate Prior: https://en.wikipedia.org/wiki/Conjugate_prior#Table_of_conjugate_distributions

式 (39)[*2]については、今回 σ を明示的に与えているので σ^2 の更新はない。もし、σ^2 もベイズにするのなら、

$$\int p(x_i|z_i = k, \mu_k, \sigma^2) p(\mu_k, \sigma_k^2|\mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \mu_P, \sigma_P^2) d\mu_k d\sigma_k^2 \tag{40}$$

のような形になるはず。

式 (38) の初項に関しては、μ_k の期待値 $(\overline{\mu_k})$ を使う近似を考えてしまって良い。

$$p(x_i|z_i = k, \mu_k, \sigma^2) \int p(\mu_k|\mathbf{x}^{\setminus i}, \mathbf{z}^{\setminus i}, \mu_P, \sigma_P^2, \sigma^2) d\mu_k \tag{41}$$

$$= p(x_i|z_i = k, \overline{\mu_k}, \sigma^2) \tag{42}$$

μ_k は色々な値を取りうるけれども、正規分布に従う。データが十分にあれば、μ_k の事後分布は充分に尖った形になるので期待値で代表させたもので近似しても良い。

[*2] これが計算しているのは、「事後分布のヒストグラム」になるはず。事後分布のヒストグラムから z を選ぶ。