

概要

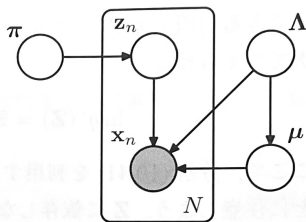
10.2 節では混合ガウス分布に変分推論法を適用する例について考えていた。変分推論は推定したい潜在変数の事後分布を推定するための手法で、サンプリング法よりも厳密でない代わりに少ない計算量で実施できる利点があった。以降では、この変分推論を用いて混合ガウス分布のパラメータである混合比やガウス分布のパラメータといった潜在変数を求める問題を解いていく。

10.2.1 節では、変分推論の一般的な結果を用いて、潜在変数の近似分布を更新する手続きを確認する。ここで注目すべき点は、潜在変数に共役事前分布を設定していることで、更新後の近似分布も同じ分布に属することである。この性質により、分布の更新式は事前分布として設定した分布のパラメータの更新式として得られる。これを確かめた後に、変分ベイズにおける更新と最尤推定におけるEMアルゴリズムを比較し、これらが類似していることを確かめる。また、この節の最後には間欠欠データに対して混合ガウス分布の変分推論を適用する例を見る。この例の挙動から、変分ベイズがEMアルゴリズムよりも優れている点について考察する。

10.2.2 節では、10.2.1 節の補足として、変分推論の実行の最中にエビデンス下界を計算できることを確かめる。エビデンス下界は、変分推論の終了条件として、収束を確かめるために利用される。

10.2.1 変分事後分布

10.2 節では混合ガウス分布に変分推論法を適用する例について考えていた。推定したい潜在変数は、混合比 π と割り当ての指示変数 \mathbf{Z} 、ガウス分布のパラメータ μ 、 Λ であった。今、混合ガウス分布からサンプル \mathbf{X} が得られるとき、変分推論により潜在変数の事後分布を得る手続きについて確かめる。グラフィカルモデルでの表現は下記のようになった。



このグラフィカルモデルからわかるように、同時分布は次のように分解することができる。

$$p(\mathbf{X}, \mathbf{Z}, \pi, \mu, \Lambda) = p(\mathbf{X} | \mathbf{Z}, \mu, \Lambda) p(\mathbf{Z} | \pi) p(\pi) p(\mu | \Lambda) p(\Lambda)$$

それぞれの因子に対応する分布は前節を参照とする。いま、潜在変数とパラメータに分解した変分近似を考える。変分推論における仮定は下記のみである。

$$q(\mathbf{Z}, \pi, \mu, \Lambda) = q(\mathbf{Z}) q(\pi, \mu, \Lambda)$$

変分推論で近似分布を求める手続きは、着目する変数を変えながら順に分布を更新していくものであった。ここでも、 $q(\mathbf{Z})$ と $q(\pi, \mu, \Lambda)$ それぞれに対して順に更新式を得る。更新で利用する式は、10.1 節で得られた一般的な更新式 (10.9) である。

$$\ln q_i^*(\mathbf{z}_i) = E_{i \setminus i} [\ln p(\mathbf{x}, \mathbf{z})] + \text{const.} \quad (10.9)$$

■ $q(\mathbf{z})$ の更新式

ここでは因子 $q(\mathbf{z})$ の更新式の導出を考える。後で確かめるように、最適な $q(\mathbf{z})$ は事前分布 $p(\mathbf{z}|\pi)$ と同じ形の分布となる。10.9 式より、

$$\begin{aligned} \ln q^*(\mathbf{z}) &= E_{\pi, \mu, \Lambda} [\ln p(\mathbf{x}, \mathbf{z}, \pi, \mu, \Lambda)] + \text{const.} \\ &= E_{\pi, \mu, \Lambda} [\ln p(\mathbf{x}|\mathbf{z}, \mu, \Lambda) + \ln p(\mathbf{z}|\pi)] + \text{const.} \\ &= E_{\mu, \Lambda} [\ln p(\mathbf{x}|\mathbf{z}, \mu, \Lambda)] + E_{\pi} [\ln p(\mathbf{z}|\pi)] + \text{const.} \\ &= E_{\mu, \Lambda} \left[\sum_n \sum_k \mathbf{z}_{nk} \ln \mathcal{N}(\mathbf{x}_n | \mu_k, \Lambda_k^{-1}) \right] \\ &\quad + E_{\pi} \left[\sum_n \sum_k \mathbf{z}_{nk} \ln \pi_k \right] + \text{const.} \quad \left(\begin{array}{l} \because (10.37) \\ \because (10.38) \end{array} \right) \\ &= \sum_n \sum_k \mathbf{z}_{nk} \left\{ E_{\mu, \Lambda} [\ln \mathcal{N}(\mathbf{x}_n | \mu_k, \Lambda_k^{-1})] + E_{\pi} [\ln \pi_k] \right\} + \text{const.} \\ &= \sum_n \sum_k \mathbf{z}_{nk} \left\{ E_{\mu, \Lambda} \left[\frac{1}{2} \ln |\Lambda_k| - \frac{D}{2} \ln(2\pi) - \frac{1}{2} (\mathbf{x}_n - \mu_k)' \Lambda_k (\mathbf{x}_n - \mu_k) \right] \right. \\ &\quad \left. + E_{\pi} [\ln \pi_k] \right\} + \text{const.} \\ &= \sum_n \sum_k \mathbf{z}_{nk} \left\{ \frac{1}{2} \ln E[|\Lambda_k|] - \frac{D}{2} \ln(2\pi) - \frac{1}{2} E_{\mu, \Lambda} [(\mathbf{x}_n - \mu_k)' \Lambda_k (\mathbf{x}_n - \mu_k)] + E_{\pi} [\ln \pi_k] \right\} + \text{const.} \\ &= \sum_n \sum_k \mathbf{z}_{nk} \ln \varphi_{nk} + \text{const.} \end{aligned}$$

$$\therefore \ln \varphi_{nk} = \frac{1}{2} \ln E[|\Lambda_k|] - \frac{D}{2} \ln(2\pi) - \frac{1}{2} E_{\mu, \Lambda} [(\mathbf{x}_n - \mu_k)' \Lambda_k (\mathbf{x}_n - \mu_k)] + E_{\pi} [\ln \pi_k]$$

両辺に指数をとると、

$$q(\mathbf{z}) \propto \prod_n \prod_k \varphi_{nk}^{\mathbf{z}_{nk}}$$

$$\therefore q(\mathbf{z}) = \prod_n \prod_k \varphi_{nk}^{\mathbf{z}_{nk}} \quad \left(\because \varphi_{nk} = \frac{\varphi_{nk}}{\sum_{k=1}^K \varphi_{nk}} \right)$$

これは、事前分布 $p(z|\pi) = \prod_{n=1}^N \prod_{k=1}^K \pi_k^{z_{nk}}$ と同じ形となっている。 $q(z)$ の分布の形状から、 $z_{nk} = 1$ となる確率は r_{nk} とできるため、期待値は次のようになる。

$$E[z_{nk}] = r_{nk}$$

このことから、量 r_{nk} は負担率を表していることになる。以降の計算の都合から次の量を定めておく。

$$N_k = \sum_{n=1}^N r_{nk} \quad (10.51)$$

$$\bar{x}_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} x_n \quad (10.52)$$

$$S_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} (x_n - \bar{x}_k)(x_n - \bar{x}_k)^T \quad (10.53)$$

$\ln q_{nk}$ は他の変数の分布のモーメントに依存するため、繰り返しで解く必要があることに注意する。

■ $q(\pi, \mu, \Lambda)$ の更新式

同様に因子 $q(\pi, \mu, \Lambda)$ の更新式の導出を考える。これは後で見ると $q(\pi)$ と $q(\mu_k, \Lambda_k)$ に分解でき、それぞれディリクレ分布とガウス-ウィシャート分布として更新式が得られる。

10.9式より、

$$\begin{aligned} \ln q^*(\pi, \mu, \Lambda) &= E_z [\ln p(x, z, \pi, \mu, \Lambda)] + \text{const.} \\ &= E_z [\ln p(x|z, \mu, \Lambda) + \ln p(z|\pi) \\ &\quad + \ln p(\pi) + \ln p(\mu|\Lambda) + \ln p(\Lambda) + \text{const.}] \\ &= E_z \left[\sum_n \sum_k z_{nk} \ln N(x_n | \mu_k, \Lambda_k^{-1}) \right] + E_z [\ln p(z|\pi)] \\ &\quad + \ln p(\pi) + \ln p(\mu|\Lambda) + \ln p(\Lambda) + \text{const.} \\ &= \underbrace{\ln p(\pi)}_{\pi \text{ のみからなる}} + E_z [\ln p(z|\pi)] \\ &\quad + \underbrace{\sum_n \sum_k E[z_{nk}] \ln N(x_n | \mu_k, \Lambda_k^{-1}) + \ln p(\mu|\Lambda) + \ln p(\Lambda)}_{\mu \text{ と } \Lambda \text{ のみからなる}} + \text{const.} \\ &= \ln q^*(\pi) + \sum_{k=1}^K q^*(\mu_k, \Lambda_k) + \text{const.} \end{aligned}$$

この分解により、 $q^*(\pi)$ と $q^*(\mu_k, \Lambda_k)$ をそれぞれ求めると、次の式を得る。

$$\begin{aligned}
 \ln q^*(\pi) &= \ln p(\pi) + E_Z[\ln p(Z|\pi)] + \text{const.} \\
 &= \ln \text{Dir}(\pi | \alpha_0) + E_Z[\ln \prod_n \prod_k \pi_k^{z_{nk}}] + \text{const.} \quad (\because (10.39), (10.37)) \\
 &= \sum_k (\alpha_0 - 1) \ln \pi_k + \sum_k \sum_n E[z_{nk}] \ln \pi_k + \text{const.} \\
 &= (\alpha_0 - 1) \sum_k \ln \pi_k + \sum_k \sum_n \underbrace{r_{nk}}_{= N_k} \ln \pi_k + \text{const.} \\
 &= (\alpha_0 + N_k - 1) \sum_k \ln \pi_k + \text{const.} \\
 &= \ln \text{Dir}(\pi | \alpha_0 + N_k)
 \end{aligned}$$

$$\therefore q^*(\pi) = \text{Dir}(\pi | \alpha_k) \quad (\because \alpha_k = \alpha_0 + N_k)$$

$$\begin{aligned}
 \ln q^*(\mu_k, \Lambda_k) &= \sum_n E[z_{nk}] \ln N(x_n | \mu_k, \Lambda_k^{-1}) + \ln p(\mu_k | \Lambda_k) + \ln p(\Lambda_k) + \text{const.} \\
 &= \sum_n r_{nk} \left\{ \frac{1}{2} \ln |\Lambda_k| - \frac{D}{2} \ln(2\pi) - \frac{1}{2} (x_n - \mu_k)' \Lambda_k (x_n - \mu_k) \right\} \\
 &\quad + \ln N(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1}) + \ln W(\Lambda_k | W_0, \nu_0) + \text{const.} \quad (\because (10.40)) \\
 &= \frac{1}{2} N_k \ln |\Lambda_k| - \frac{1}{2} \sum_n r_{nk} (x_n - \mu_k)' \Lambda_k (x_n - \mu_k) \\
 &\quad + \left\{ \frac{1}{2} \ln |\beta_0 \Lambda_k| - \frac{1}{2} (\mu_k - m_0)' (\beta_0 \Lambda_k) (\mu_k - m_0) \right\} \\
 &\quad + \frac{\nu_0 - D - 1}{2} \ln |\Lambda_k| - \frac{1}{2} \text{Tr}(W_0^{-1} \Lambda_k) + \text{const.}
 \end{aligned}$$

ここで、天下りではあるが $q^*(\mu_k, \Lambda_k)$ がガウス-ウィシャート分布となることを見越して、 μ_k と Λ_k に関わる項と、 Λ_k のみに関わる項に分解して考える。
 μ_k と Λ_k に関わる項については、

$$\begin{aligned}
\ln Q^*(\mu_k | \Lambda_k) &= -\frac{1}{2} \sum_n r_{nk} \mu_k' \Lambda_k \mu_k + \sum_n r_{nk} x_n' \Lambda_k \mu_k \\
&\quad - \frac{\beta_0}{2} \mu_k' \Lambda_k \mu_k + \beta_0 m_0' \Lambda_k \mu_k \\
&= \underbrace{-\frac{1}{2} \mu_k' (N_k + \beta_0) \Lambda_k \mu_k}_{\mu_k \text{ の 2 次の項}} + \underbrace{(N_k \bar{x}_k + \beta_0 m_0)' \Lambda_k \mu_k}_{\mu_k \text{ の 1 次の項}} + \text{const.}
\end{aligned}$$

$$= \ln N(\mu_k | m_k, \beta_k \Lambda_k)$$

$$\text{したがって, } \begin{cases} \beta_k = \beta_0 + N_k \\ m_k = \frac{1}{\beta_k} (\beta_0 m_0 + N_k \bar{x}_k) \end{cases}$$

Λ_k のみに関わる項については、 $Q^*(\mu_k | \Lambda_k)$ に現れる Λ_k の項も含めて考慮する必要がある、次のように求まる。

$$\begin{aligned}
\ln Q^*(\Lambda_k) &= \ln Q^*(\mu_k, \Lambda_k) - \ln Q^*(\mu_k | \Lambda_k) \\
&= \frac{1}{2} N_k \ln |\Lambda_k| - \frac{1}{2} \sum_n r_{nk} (x_n - \mu_k)' \Lambda_k (x_n - \mu_k) \\
&\quad + \left\{ \frac{1}{2} \ln |\beta_0 \Lambda_k| - \frac{1}{2} (\mu_k - m_0)' (\beta_0 \Lambda_k) (\mu_k - m_0) \right\} \\
&\quad + \frac{V_0 - D - 1}{2} \ln |\Lambda_k| - \frac{1}{2} \text{Tr} (W_0^{-1} \Lambda_k) \\
&\quad - \left\{ \frac{1}{2} \ln |\beta_k \Lambda_k| - \frac{1}{2} (\mu_k - m_k)' (\beta_k \Lambda_k) (\mu_k - m_k) \right\} \\
&\quad + \text{const.}
\end{aligned}$$

$\left. \begin{aligned} &\ln Q^*(\mu_k, \Lambda_k) \\ &-\ln Q^*(\mu_k | \Lambda_k) \end{aligned} \right\}$

$$\begin{aligned}
&= \frac{1}{2} N_k \ln |\Lambda_k| + \frac{1}{2} \left\{ \underbrace{\ln |\beta_0|}_{\rightarrow \text{const.}} + \ln |\Lambda_k| \right\} + \frac{V_0 - D - 1}{2} \ln |\Lambda_k| - \frac{1}{2} \left\{ \underbrace{\ln |\beta_k| + \ln |\Lambda_k|}_{\rightarrow \text{const.}} \right\} \\
&\quad - \frac{1}{2} \sum_n r_{nk} (x_n - \mu_k)' \Lambda_k (x_n - \mu_k) - \frac{1}{2} (\mu_k - m_0)' (\beta_0 \Lambda_k) (\mu_k - m_0) \\
&\quad - \frac{1}{2} \text{Tr} (W_0^{-1} \Lambda_k) + \frac{1}{2} (\mu_k - m_k)' (\beta_k \Lambda_k) (\mu_k - m_k) + \text{const.} \\
&= \frac{V_0 + N_k - D - 1}{2} \ln |\Lambda_k| \\
&\quad - \frac{1}{2} \text{Tr} \left\{ \underbrace{\Lambda_k \sum_n r_{nk} (x_n - \mu_k) (x_n - \mu_k)'}_{\text{赤下線部}} - \frac{1}{2} \text{Tr} \left\{ \underbrace{\beta_0 \Lambda_k (\mu_k - m_0) (\mu_k - m_0)'}_{\text{赤下線部}} \right\} \right\} \\
&\quad - \frac{1}{2} \text{Tr} \left\{ \underbrace{\Lambda_k W_0^{-1}}_{\text{赤下線部}} \right\} + \frac{1}{2} \text{Tr} \left\{ \underbrace{\beta_k \Lambda_k (\mu_k - m_k) (\mu_k - m_k)'}_{\text{赤下線部}} \right\} + \text{const.}
\end{aligned}$$

赤下線部の原理はわからないが
こうするとうまくいく。

$$\begin{aligned}
&= \frac{V_0 + N_k - D - 1}{2} \ln |\Lambda_k| \\
&\quad - \frac{1}{2} \text{Tr} \left[\Lambda_k \left\{ \sum_n \Gamma_{nk} (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)' + \beta_0 (\mu_k - m_0)(\mu_k - m_0)' \right. \right. \\
&\quad \quad \left. \left. + W_0^{-1} - \beta_k (\mu_k - m_k)(\mu_k - m_k)' \right\} \right] + \text{const.} \\
&= \frac{V_0 + N_k - D - 1}{2} \ln |\Lambda_k| \\
&\quad - \frac{1}{2} \text{Tr} \left[\Lambda_k \left\{ N_k S_k + \beta_0 (\mu_k - m_0)(\mu_k - m_0)' + W_0^{-1} \right. \right. \\
&\quad \quad \left. \left. - (\beta_0 + N_k) \left(\mu_k - \frac{1}{\beta_0 + N_k} (\beta_0 m_0 + N_k \bar{x}_k) \right) \left(\mu_k - \frac{1}{\beta_0 + N_k} (\beta_0 m_0 + N_k \bar{x}_k) \right)' \right\} \right] + \text{const.} \\
&= (\text{この値が式かわからず}) \\
&= \frac{V_0 + N_k - D - 1}{2} \ln |\Lambda_k| \\
&\quad - \frac{1}{2} \text{Tr} \left[\Lambda_k \left\{ N_k S_k + W_0^{-1} + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{x}_k - m_0)(\bar{x}_k - m_0)' \right\} \right] \\
&= \frac{V_k - D - 1}{2} \ln |\Lambda_k| - \frac{1}{2} \text{Tr} (\Lambda_k W_k^{-1}) + \text{const.} \\
&= \ln \mathcal{W}(\Lambda_k | W_k, V_k)
\end{aligned}$$

ここで、 $V_k = V_0 + N_k$

$$W_k^{-1} = N_k S_k + W_0^{-1} + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{x}_k - m_0)(\bar{x}_k - m_0)'$$

以上より、 $q^*(\mu_k, \Lambda_k)$ 全体は下記のガウス-ウィシャート分布が得られる。

$$q^*(\mu_k, \Lambda_k) = \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)') \mathcal{W}(\Lambda_k | W_k, V_k)$$

前半の $q(z)$ の更新式に含まれる期待値は、ここまでで得られたディリクレ分布とガウスウィシャート分布を用いて計算される。

$$q(z) \propto \prod_n \prod_k \varphi_{nk}^{z_{nk}}$$

$$\text{ここで、} \ln \varphi_{nk} = \underbrace{E[\ln \pi_k]} + \frac{1}{2} \underbrace{E[\ln |\Lambda_k|]} - \frac{D}{2} \ln(2\pi) - \frac{1}{2} \underbrace{E_{\mu_k, \Lambda_k}[(\mathbf{x}_n - \mu_k)' \Lambda_k^{-1} (\mathbf{x}_n - \mu_k)]} \quad (10.64)$$

赤下線部は次のように計算される。(証明略)

$$E[\ln \pi_k] = \psi(\alpha_k) - \psi(\hat{\alpha}) \equiv \ln \tilde{\Lambda}_k \quad (10.66)$$

$$E[\ln |\Lambda_k|] = \sum_{i=1}^D \psi\left(\frac{V_k - 1 - i}{2}\right) + D \ln 2 + \ln |W_k| \equiv \ln \tilde{W}_k \quad (10.65)$$

$$E_{\mu_k, \Lambda_k}[(\mathbf{x}_n - \mu_k)' \Lambda_k^{-1} (\mathbf{x}_n - \mu_k)] = D \beta_k^{-1} + V_k (\mathbf{x}_n - m_k)' W_k (\mathbf{x}_n - m_k) \quad (10.64)$$

ここで ψ はディリクレ関数である。

ここまでで変分推論における全ての更新式を導出することができた。全体を振り返ると、潜在変数の同時分布を次のように分解し、それぞれの因子を更新していくのであった。

$$q(z, \pi, \mu, \Lambda) = q(z) q(\pi, \mu, \Lambda)$$

この更新は最尤推定におけるEMアルゴリズムに似ている。というのも、 $q(z)$ を更新する操作は、負担率 $E[z_{nk}] = r_{nk}$ を計算すること(9.23式)に似ており、 $q(\pi, \mu, \Lambda)$ を更新する操作は負担率を利用してパラメータを更新すること(9.24-26式)に似ている。

このことをより深く理解するために、両者における負担率の計算式を比べてみる。変分ベイズ法における負担率の式は、(10.46), (10.64), (10.65), (10.66) を (10.49) に代入して、次のようになる。

$$\begin{aligned} r_{nk} &\propto \varrho_{nk} = \exp \left\{ E[\ln \pi_k] + \frac{1}{2} E[\ln |\Lambda_k|] - \frac{D}{2} \ln(2\pi) - \frac{1}{2} E_{\mu_k, \Lambda_k} [(x_n - \mu_k)' \Lambda_k^{-1} (x_n - \mu_k)] \right\} \\ &= \exp \left\{ \ln \tilde{\pi}_k + \frac{1}{2} \ln |\tilde{\Lambda}_k| - \frac{D}{2 \ell_k} - \frac{V_k}{2} (x_n - m_k)' W_k (x_n - m_k) \right\} \\ &= \tilde{\pi}_k |\tilde{\Lambda}_k|^{\frac{1}{2}} \exp \left\{ -\frac{V_k}{2 \ell_k} - \frac{V_k}{2} (x_n - m_k)' W_k (x_n - m_k) \right\} \quad (10.67) \end{aligned}$$

一方で、EMアルゴリズムでの負担率の計算は(9.13) 式より次のようになり、上記 (10.67) 式と類似した式となっている。

$$\begin{aligned} r_{nk} &\propto \pi_k \mathcal{N}(x | \mu_k, \Lambda_k^{-1}) \\ &= \pi_k \frac{1}{(2\pi)^{D/2}} |\Lambda_k|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)' \Lambda_k^{-1} (x - \mu_k) \right\} \\ &\propto \pi_k |\Lambda_k|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)' \Lambda_k^{-1} (x - \mu_k) \right\} \quad (10.68) \end{aligned}$$

これらの類似性から、変分ベイズにおいて $q(z)$ を更新する操作は変分Eステップと呼ばれ、 $q(\pi, \mu, \Lambda)$ を更新する操作は変分Mステップと呼ばれる。

ここまでで混合ガウス分布に対する変分ベイズ法の手続きを見てきたが、実際の間欠泉データを使用してモデルの特徴について考察してみる。特に、EMアルゴリズムと比較して、特定のデータにガウス分布の要素が当てはまるような問題が起きないことや、混合要素 K を増やしても過学習しにくいといったことが利点として挙げられる。

混合要素は $K=6$ とした混合ガウス分布で、事前分布は十分になだらかなものを使用した際に、変分ベイズ法の各繰り返しごとのモデルは次の図のようになる。

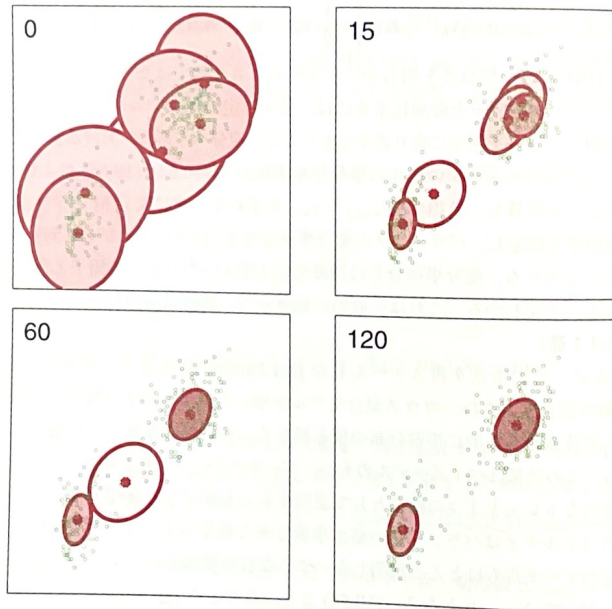


図 10.6 Old Faithful 間欠泉データ集合に対して、 $K = 6$ 混合の変分ベイズ混合ガウスモデルを適用した結果。各混合要素について、楕円は確率密度が標準偏差の 1 倍となる等高線を表し、楕円内の赤の濃さはその混合要素の持つ混合係数の期待値に比例している。各図の左上の数字は変分推論の繰り返しの回数を表している。混合係数の期待値が 0 と数値的に区別できなくなった混合要素については、プロットを省略した。

(この図の解説は細かい話題が複数記載されているので、p.193中段から読み合わせとしたい)
重要な点は、p.194 下段における内容である、混合ガウス分布の要素が特定のデータを説明するような「縮退」が起こらないことと、混合要素数 K に大きな値を選んだ場合でも複雑なモデルを避けるベイズの特性から過学習が起こらないという点である。

10.2.2 変分下界

変分ベイズは繰り返し法により事後分布の近似を行うが、これの収束判定にエビデンス下界 (ELBO) を観察することは有用である。(教科書ではエビデンス下界という用語が出ないが、「ベイズ深層学習」p.94 で説明されるエビデンス下界と同一の概念なのでわかりやすさのためにこう呼びます)

変分推論における各繰り返しでエビデンス下界は決して減少しないため、ソフトウェアの実装の正しさの検証にも利用できる。エビデンス下界が減少しないことは、近似分布の更新式 (10.9) が、下界の変分により常に停留点となるように導き出されたことから理解できる。

混合ガウス分布における下界は (10.3) 式を変形して次のように求まる。

$$\begin{aligned}
 L &= \sum_z \iiint q(z, \pi, \mu, \Lambda) \ln \left\{ \frac{p(x, z, \pi, \mu, \Lambda)}{q(z, \pi, \mu, \Lambda)} \right\} d\pi d\mu d\Lambda \\
 &= E_q [\ln p(x, z, \pi, \mu, \Lambda)] - E_q [\ln q(z, \pi, \mu, \Lambda)] \\
 &= E_q [\ln p(x|z, \mu, \Lambda)] + E_q [\ln p(z|\pi)] + E_q [\ln p(\pi)] + E_q [\ln p(\mu, \Lambda)] \\
 &\quad - E_q [\ln q(z)] - E_q [\ln q(\pi)] - E_q [\ln q(\mu, \Lambda)]
 \end{aligned}$$

これらの各項については次のように求めることができる。

$$\begin{aligned}
 E[\ln p(x|z, \mu, \Lambda)] &= E[\ln \{ \prod_k \pi_k \mathcal{N}(x_k | \mu_k, \Lambda_k^{-1})^{z_{nk}} \}] \quad (\because (10.38)) \\
 &= E \left[\sum_n \sum_k z_{nk} \left\{ -\frac{D}{2} \ln(2\pi) + \ln |\Lambda_k| - \frac{1}{2} (x_n - \mu_k)' \Lambda_k (x_n - \mu_k) \right\} \right] \\
 &= \frac{1}{2} \sum_n \sum_k E(z_{nk}) \left\{ -D \ln(2\pi) + E[\ln |\Lambda_k|] - E[(x_n - \mu_k)' \Lambda_k (x_n - \mu_k)] \right\} \\
 &= \frac{1}{2} \sum_n \sum_k \underbrace{r_{nk}}_{(10.50)} \left\{ -D \ln(2\pi) + \underbrace{\ln \tilde{\Lambda}_k}_{(10.65)} - \underbrace{\frac{D}{\tilde{\theta}_k} - \mathcal{V}_k (x_n - m_k)' \mathcal{W}_k (x_n - m_k)}_{(10.64)} \right\} \quad (10.71)
 \end{aligned}$$

$$\begin{aligned}
 E[\ln p(z|\pi)] &= E[\ln \{ \prod_k \pi_k \pi_k^{z_{nk}} \}] \quad (\because (10.39)) \\
 &= E \left[\sum_n \sum_k z_{nk} \ln(\pi_k) \right] \\
 &= \sum_n \sum_k E[z_{nk}] \cdot E[\ln(\pi_k)] \\
 &= \sum_n \sum_k \underbrace{r_{nk}}_{(10.50)} \cdot \ln \tilde{\pi}_k \quad (\because (10.50), (10.66)) \quad (10.72)
 \end{aligned}$$

$$\begin{aligned}
E[\ln p(\pi)] &= E\left[\ln \left\{ \text{Dir}(\pi | \alpha_0) \right\}\right] \\
&= E\left[\ln C(\alpha_0) + (\alpha_0 - 1) \sum_k \ln \pi_k\right] \\
&= \ln C(\alpha_0) + (\alpha_0 - 1) \sum_k \ln \tilde{\pi}_k \quad (\because (10.66)) \quad (10.73)
\end{aligned}$$

$$\begin{aligned}
E[\ln p(\mu, \Lambda)] &= E\left[\ln \left\{ \prod_k N(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1}) \mathcal{W}(\Lambda_k | W_0, \nu_0) \right\}\right] \\
&= \langle \text{展開して} E \text{を適用する} \rangle \\
&= \frac{1}{2} \sum_k \left\{ D \ln \left(\frac{\beta_0}{2\pi} \right) + \ln \tilde{\Lambda}_k - \frac{D \beta_0}{\beta_k} - \beta_0 \nu_k (m_k - m_0)' (W_k (m_k - m_0)) \right\} \\
&\quad + K \ln B(W_0, \nu_0) + \frac{\nu_0 - D - 1}{2} \sum_k \ln \tilde{\Lambda}_k - \frac{1}{2} \sum_k \nu_k \text{Tr} (W_0^{-1} W_k) \quad (10.74)
\end{aligned}$$

$$\begin{aligned}
E[\ln q(Z)] &= E\left[\ln \left\{ \prod_k r_{nk} z_{nk} \right\}\right] \quad (\because (10.48)) \\
&= E\left[\sum_n \sum_k z_{nk} \ln(r_{nk})\right] \\
&= \sum_n \sum_k r_{nk} \ln(r_{nk}) \quad (10.75)
\end{aligned}$$

$$\begin{aligned}
E[\ln q(\pi)] &= E\left[\ln \left\{ \text{Dir}(\pi | \alpha) \right\}\right] \quad (\because (10.59)) \\
&= E\left[\ln \left\{ C(\alpha) \prod_k \pi_k^{\alpha_k - 1} \right\}\right] \\
&= \sum_k (\alpha_k - 1) E(\ln \pi_k) + \ln C(\alpha) \\
&= \sum_k (\alpha_k - 1) \ln \tilde{\pi}_k + \ln C(\alpha) \quad (\because (10.66)) \quad (10.76)
\end{aligned}$$

$$\begin{aligned}
E[\ln q(\mu, \Lambda)] &= E\left[\ln \left\{ \prod_k N(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) \mathcal{W}(\Lambda_k | W_k, \nu_k) \right\}\right] \\
&= \langle \text{展開して} E \text{を適用する} \rangle \\
&= \sum_k \left\{ \frac{1}{2} \ln \tilde{\Lambda}_k + \frac{D}{2} \ln \left(\frac{\beta_k}{2\pi} \right) - \frac{D}{2} - H[q(\Lambda_k)] \right\}
\end{aligned}$$

以上より、これまで得られた式を使ってエビデンス下界を算出できることを示せた。