1　例：変分混合ガウス分布

前回は変分推論の枠組みの中でモデル分布の事後分布を求めた。今回はその事後分布を用いて予測分布を求めたり、また事後分布そのものについて実用的な側面から考察を加えたりする。

さらに変分理論の 2 つ目の例としてベイズ線形回帰モデルの構築を考える。実は変分理論によるモデルの解は、エビデンス最大化によるものと同じなることを確かめることができる。

1.1　予測分布

前回求めた事後分布をもとに、潜在変数およびパラメタについての畳み込み積分によって、新しい観測値 $\hat{\boldsymbol{x}}$ の予測分布を求めよう。

$$p(\hat{\boldsymbol{x}} \mid \boldsymbol{X}) = \sum_{\hat{\boldsymbol{z}}} \int\int\int p(\hat{\boldsymbol{x}} \mid \hat{\boldsymbol{z}}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\hat{\boldsymbol{z}} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \boldsymbol{X}) d\boldsymbol{\pi} d\boldsymbol{\mu} d\boldsymbol{\Lambda} \tag{1}$$

式 1 において、$\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}$ について条件付き分布になっている部分に対して、それぞれ式（10.37）と（10.38）を用いる。すると式 1 は次のように書ける。あとで手を動かす

$$p(\hat{\boldsymbol{x}} \mid \boldsymbol{X}) = \tag{2}$$

この積分は解析的に実行することができない。そこで、パラメタについての真の事後分布 $p(\hat{\boldsymbol{x}} \mid \hat{\boldsymbol{z}}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ を変分事後分布 $q(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = q(\boldsymbol{\pi}) \prod_{k=1}^{K} q(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ に置き換える。これによって真の予測分布は次のようになる。混合 t 分布の導出は諦めた。

$$p(\hat{\boldsymbol{x}} \mid \boldsymbol{X}) \simeq \tag{3}$$

データ数 N が大きいとき式 3 は混合ガウス分布に近づく。この極限計算も諦めた。

1.2　混合要素数の決定

このサブセクションは読み合わせ。

1.3　導出された分解

このサブセクションは読み合わせ。

2　変分線形回帰

変分推論の活用事例その 2 として、ベイズ線形回帰モデルを本枠組みで構築することを考える。ベイズ線形回帰モデルは以前エビデンス最大化の枠組みで構築を行なったものであり、今回はそれぞれの枠組みで得られる解が一致することを確かめる。

ここでエビデンス最大化の枠組みを簡単に振り返りたい。上巻 p.XXX を読み合わせ。どんな式を定義したのか、どんな近似を適用したのか思い出すことにしよう。

それでは変分推論の枠組みで、ベイズ線形回帰モデルを構築していく。議論を簡単にするため、ノイズの精度パラメタ β は既知かつ固定されたものとする。

重みおよびその分散パラメタに関する事後分布を求めたい。そのために準備として、条件付き分布と事前分布を式（10.87-89）のように定義する。またこれらの定義から全ての変数 $\boldsymbol{w}, \alpha, t$ についての同時分布は式（10.90）のようになる。そして、式（10.87-90）は有向グラフを使って、図 10.8 のように表すことができる。

2.1 変分分布

まずは後の議論のベースとなる事後分布 $p(\boldsymbol{w}, \alpha \mid \boldsymbol{t})$ について、そのモデル分布 $q(\boldsymbol{w}, \alpha)$ を求めよう。モデル分布は平均場近似によって

$$q(\boldsymbol{w}, \alpha) = q(\boldsymbol{w})q(\alpha) \tag{4}$$

と仮定することができる。

次に、式（10.9）を用いて各因子を計算する。まず $q(\alpha)$ については

$$\begin{aligned}
\ln q^*(\alpha) &= E_{\boldsymbol{w}}[\ln p(\boldsymbol{w}, \alpha)] + const \\
&= E_{\boldsymbol{w}}[\ln p(\alpha)(\boldsymbol{w} \mid \alpha)] + const \\
&= E_{\boldsymbol{w}}[\ln p(\alpha) + \ln p(\boldsymbol{w} \mid \alpha)] + const \\
&= \ln p(\alpha) + E_{\boldsymbol{w}}[\ln p(\boldsymbol{w} \mid \alpha)] + const \\
&= \ln Gam(\alpha \mid a_0, b_0) + E_{\boldsymbol{w}}[\ln \mathcal{N}(\boldsymbol{w} \mid \boldsymbol{0}, \alpha^{-1}\boldsymbol{I})] + const \\
&= \ln \frac{1}{\Gamma(a_0)} b_0^{a_0} \alpha^{a_0 - 1} e^{-b_0\alpha} + E_{\boldsymbol{w}}[\ln (\frac{\alpha}{2\pi})^{\frac{M}{2}} \exp(-\frac{\alpha \boldsymbol{w}^T\boldsymbol{w}}{2})] + const \\
&= \ln \alpha^{a_0 - 1} e^{-b_0\alpha} + E_{\boldsymbol{w}}[\frac{M}{2} \ln \frac{\alpha}{2\pi} - \frac{\alpha \boldsymbol{w}^T\boldsymbol{w}}{2}] + const \\
&= (a_0 - 1)\ln \alpha - b_0\alpha + \frac{M}{2}\ln \alpha - \frac{\alpha}{2} E_{\boldsymbol{w}}[\boldsymbol{w}^T\boldsymbol{w}] + const \\
&= (a_0 + \frac{M}{2} - 1)\ln \alpha - (b_0 + \frac{1}{2} E_{\boldsymbol{w}}[\boldsymbol{w}^T\boldsymbol{w}])\alpha + const
\end{aligned} \tag{5}$$

となる。結局 $q^*(\alpha)$ はガンマ分布[*1]で与えられ、その分布のパラメタは式（10.94-95）のようになる。そして $q^*(\alpha)$ を求めるには重みについての二次モーメントが必要になることがわかる。

同様にして、$q(\boldsymbol{w})$ についても考える。

$$\begin{aligned}
\ln q^*(\boldsymbol{w}) &= E_\alpha[\ln p(\boldsymbol{t}, \boldsymbol{w}, \alpha)] + const \\
&= E_\alpha[\ln p(\boldsymbol{t} \mid \boldsymbol{w}) p(\boldsymbol{w} \mid \alpha) p(\alpha)] + const \\
&= E_\alpha[\ln p(\boldsymbol{t} \mid \boldsymbol{w}) + \ln p(\boldsymbol{w} \mid \alpha) + \ln p(\alpha)] + const \\
&= E_\alpha[\ln \prod_{n=1}^{N} \mathcal{N}(t_n \mid \boldsymbol{w}^T\boldsymbol{\phi}_n, \beta^{-1}) + \ln \mathcal{N}(\boldsymbol{w} \mid \boldsymbol{0}, \alpha^{-1}\boldsymbol{I})] + const \\
&= E_\alpha[\sum_{n=1}^{N} \ln \mathcal{N}(t_n \mid \boldsymbol{w}^T\boldsymbol{\phi}_n, \beta^{-1}) + \frac{M}{2}\ln \frac{\alpha}{2\pi} - \frac{\alpha \boldsymbol{w}^T\boldsymbol{w}}{2}] + const \\
&= \sum_{n=1}^{N} \ln [\frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp(-\frac{\beta}{2}(t_n - \boldsymbol{w}^T\boldsymbol{\phi}_n)^2)] - \frac{\boldsymbol{w}^T\boldsymbol{w}}{2} E_\alpha(\alpha) + const \\
&= -\frac{\beta}{2} \sum_{n=1} N(t_n - \boldsymbol{w}^T\boldsymbol{\phi}_n)^2 - \frac{1}{2} E_\alpha(\alpha)\boldsymbol{w}^T\boldsymbol{w} + const
\end{aligned} \tag{6}$$

[*1] 式 5 最右辺の第一項係数が（B.26）の a に相当し、第二項係数が b に相当する。

$q^*(\boldsymbol{w})$ を求めるにはパラメタについての一次モーメントが必要になる。ここから $q^*(\boldsymbol{w})$ を平方完成すれば、式（10.99-101）が導かれる[*2]。つまり $q^*(\boldsymbol{w})$ はガウス分布で与えられる。

モーメントの計算および再推定の話は読み合わせとする。さらにエビデンス最大化による解との比較についても読み合わせとする。

2.2 予測分布

先のサブセクションで求めた事後分布をもとに、重みについての畳み込み積分を用いることで、新しい $\boldsymbol{x}$ に対する予測分布は次のようになる。

$$\begin{aligned}
p(t \mid \boldsymbol{x}, \boldsymbol{t}) &= \int p(t \mid \boldsymbol{x}, \boldsymbol{w}) p(\boldsymbol{w} \mid \boldsymbol{t}) d\boldsymbol{w} \\
&\simeq \int p(t \mid \boldsymbol{x}, \boldsymbol{w}) q(\boldsymbol{w}) d\boldsymbol{w} \\
&= \int \mathcal{N}(t \mid \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}), \beta^{-1})(\boldsymbol{w} \mid \boldsymbol{m}_N, \boldsymbol{S}_N) d\boldsymbol{w} \\
&= \mathcal{N}(t \mid \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{m}_N + 0, \frac{1}{\beta} + \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{S}_N \phi(\boldsymbol{x})) \ (\because \ (2.115)) \\
&= \mathcal{N}(t \mid \boldsymbol{m}_N^T \boldsymbol{\phi}(\boldsymbol{x}), \sigma(\boldsymbol{x})^2)
\end{aligned} \tag{7}$$

なお $\sigma(\boldsymbol{x})^2$ は式（10.106）で定義される。これは式（3.59）と同様の結果が得られている[*3]。

[*2] 平方完成の計算そのものは冗長であるため、今回はスキップする。

[*3] ただし α が $E(\alpha)$ に置き換わっている。