混合正規分布に対する変分ベイズ法

東京工業大学　大橋耕也

- ベイズ統計は「確率モデルを事後分布により平均をして、予測分布を計算し、それはおおよそ真の分布に近いだろう」と推定する枠組みです。

$$(\text{予測分布}) = \mathbb{E}_w[p(x|w)]$$

- しかし、$\mathbb{E}_w[(\ )]$を簡単に計算できない場合も当然ありますし、むしろその方が普通かと思います。

- 「ベイズ統計の理論と方法」第５章には、事後分布を近似的に計算する枠組みが紹介されています。（MCMC法と変分ベイズ法）

- 今回はこの本でも紹介されている、混合正規分布に対する変分ベイズ法を実験してみました。結果だけを述べるので、詳しくは本を参考にしてください。

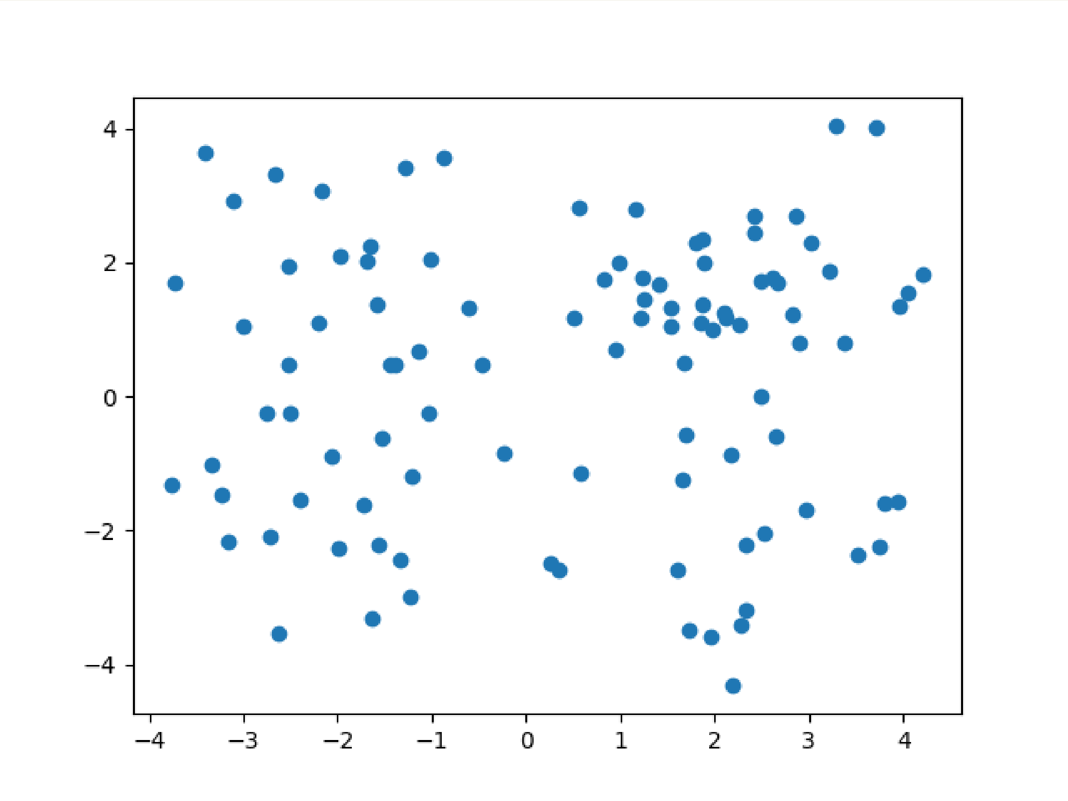
平均場近似

- 同時確率密度関数 $p(w,y)$ を確率密度関数 $q(w), r(y)$ の積で近似する手法のことです。
- 近似は、KL情報量を最小にすることで行います。

$$q(w)r(y) = \underset{qr}{\operatorname{argmin}}\, KL(q(w)r(y)\ |\, p(w,y))$$

- 事後分布を平均場近似により近似をする手法を特に変分ベイズ法といい、特に指数型分布の混合分布に対しては反復法による予測分布計算法を与える為相性がいいです。

混合正規分布に対する変分ベイズ法



- 左のようなデータが目の前にあるとき、「何個のグループに分けられますか？」という質問にどう答えられるでしょうか

- 混合正規分布

$$p(x|w) = \sum a_j N(m_j, S)$$

によるモデル化はそういった質問に答える為の一つの手立てです。

混合正規分布に対する変分ベイズ法

確率モデル

$$p(x|w) = \sum_{k=1}^{K} a_k N(x|m_k, S)$$

事前分布

$$\begin{aligned} \varphi(w) &= \varphi_1(a) * \varphi_2(m) \\ &= Dir(a|\phi) * Con(m|\eta_1, \eta_2) \end{aligned}$$

事前分布は、各正規分布の重みに対してはディリクレ分布、各正規分布の平均に対しては共役事前分布（指数型分布）を用います

混合正規分布に対する変分ベイズ法

- 以下の反復によりハイパーパラメータ(ϕ,η)が求まります
- (1)初期化

$$\hat{\phi}_k = \frac{n}{K} + \phi_k, \hat{\eta}_{k1} = \sum_k x_i + \eta_{k1} + noise, \hat{\eta}_{k2} = \frac{n}{K} + \eta_{k2} + noise$$

- (2)隠れ変数yの更新

$$\hat{y}_{ik} \leftarrow \frac{\exp(L_{ik})}{\sum_k \exp(L_{ik})}$$

$$L_{ik} = \psi\left(\hat{\phi}_k\right) - \psi\left(n + \sum_k \hat{\phi}_k\right) - \frac{\dim(x)}{2\,\hat{\eta}_{k2}} - \frac{1}{2}\left(\frac{\hat{\eta}_{k1}}{\hat{\eta}_{k2}} - x_i\right)^{\top} S^{-1}\left(\frac{\hat{\eta}_{k1}}{\hat{\eta}_{k2}} - x_i\right)$$

ψ：ディガンマ関数

- (3)ハイパーパラメータの更新

$$\hat{\phi}_k \leftarrow \sum_i \hat{y}_{ik} + \phi_k, \hat{\eta}_{k1} \leftarrow \sum_i \hat{y}_{ik} x_i + \eta_{k1}, \hat{\eta}_{k2} \leftarrow \sum_i \hat{y}_{ik} + \eta_{k2}$$

- (2),(3)を繰り返す

混合正規分布に対する変分ベイズ法

- 反復法により得られたハイパーパラメータ$(\hat{\phi}, \hat{\eta})$により事前分布が

$$\varphi(w) = Dir\left(a \middle| \hat{\phi}\right) * Con(m|\hat{\eta})$$

と定まり、予測分布も以下の式で与えられます。

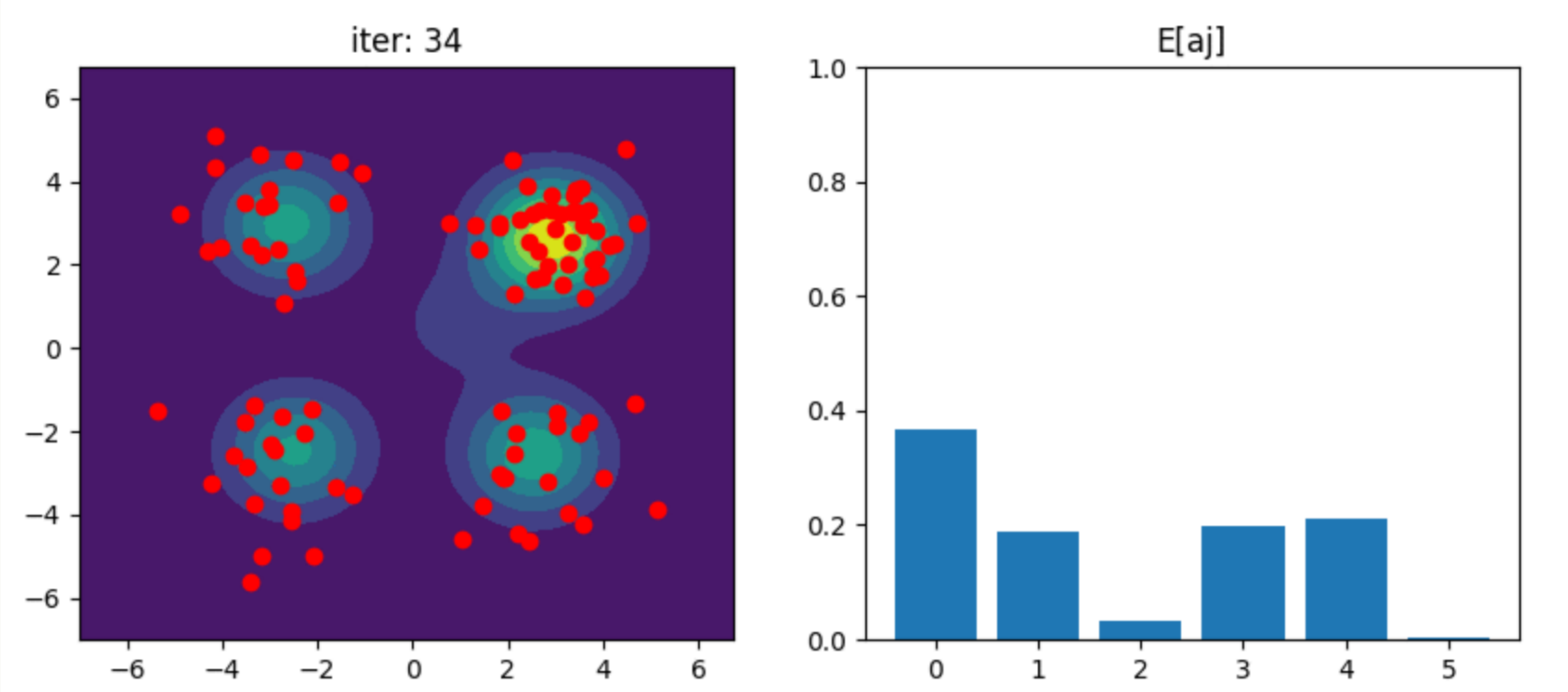
$$p^*(x) = \sum_{k=1}^{K} \left(\frac{\hat{\phi}_k}{\sum_k \hat{\phi}_k} \right) N(x|0, S) \frac{z(\hat{\eta}_k + g(x))}{z(\hat{\eta}_k)}$$

ただし、

$$z(\eta_k) = \left(N(\eta_{k1}|0, \eta_{k2} S) * \eta_{k2}^{-M} \right)^{-1}$$

■ 4つのグループから合計データを100個生成し、混合正規分布によりモデリングをした結果は以下のようになりました。

（真のコンポーネント数４、K=6, S=単位行列）



- 各グループ間の距離を分散Sと同程度に小さくした場合は、推定がうまくいっていないこともわかりました

(真のコンポーネント数:4, K=6, S=単位行列)