

Variational Inference GMM

「ベイズ推論による機械学習」（須山敦志著）の4章のガウス混合モデルを参考にした。事後分布の推論には変分推論を使った。予測分布とクラス予測平均の計算は詳しく記載されていなかったなので、以下のように導出して実装した。

予測分布の導出

モデルの定義は書籍の4.4節を参照。導出は4.4.4節の崩壊型ギブスサンプリングを参考にした。

予測分布 $p(\mathbf{x}_* | \mathbf{X}, \mathbf{S})$ の導出

$$\begin{aligned} P(\mathbf{x}_* | \mathbf{X}, \mathbf{S}) &= \sum_{k=1}^K p(\mathbf{x}_*, s_{*,k} = 1 | \mathbf{X}, \mathbf{S}) \\ &= \sum_{k=1}^K p(\mathbf{x}_* | s_{*,k} = 1, \mathbf{X}, \mathbf{S}) p(s_{*,k} = 1 | \mathbf{X}, \mathbf{S}) \\ &= \sum_{k=1}^K \text{St} \left(\mathbf{x}_* \middle| \mathbf{m}_k, \frac{(1 - D + \hat{\nu}_k) \hat{\beta}_k}{1 + \hat{\beta}_k} \hat{\mathbf{W}}_k, 1 - D + \hat{\nu}_k \right) \text{Cat} \left(s_{*,k} = 1 \middle| \left(\frac{\hat{\alpha}_k}{\sum_{i=1}^K \hat{\alpha}_i} \right)_{k=1}^K \right) \\ &= \sum_{k=1}^K \frac{\hat{\alpha}_k}{\sum_{i=1}^K \hat{\alpha}_i} \text{St} \left(\mathbf{x}_* \middle| \mathbf{m}_k, \frac{(1 - D + \hat{\nu}_k) \hat{\beta}_k}{1 + \hat{\beta}_k} \hat{\mathbf{W}}_k, 1 - D + \hat{\nu}_k \right) \\ &\quad \left(\text{s.t.} \quad \sum_{k=1}^K \frac{\hat{\alpha}_k}{\sum_{i=1}^K \hat{\alpha}_i} = 1 \right) \end{aligned}$$

クラス予測平均 $\langle \mathbf{s}_* \rangle_{p(\mathbf{s}_* | \mathbf{x}_*, \mathbf{X}, \mathbf{S})}$ の導出

$$\begin{aligned} \langle \mathbf{s}_* \rangle_{p(\mathbf{s}_* | \mathbf{x}_*, \mathbf{X}, \mathbf{S})} &= \sum_{k=1}^K \mathbf{e}_k p(s_{*,k} = 1 | \mathbf{x}_*, \mathbf{X}, \mathbf{S}) \\ &\propto \sum_{k=1}^K \mathbf{e}_k p(\mathbf{x}_* | s_{*,k} = 1, \mathbf{X}, \mathbf{S}) p(s_{*,k} = 1 | \mathbf{X}, \mathbf{S}) \\ &= \left(\frac{\hat{\alpha}_k}{\sum_{i=1}^K \hat{\alpha}_i} \text{St} \left(\mathbf{x}_* \middle| \mathbf{m}_k, \frac{(1 - D + \hat{\nu}_k) \hat{\beta}_k}{1 + \hat{\beta}_k} \hat{\mathbf{W}}_k, 1 - D + \hat{\nu}_k \right) \right)_{k=1}^K \\ &\quad \left(\text{s.t.} \quad \sum_{k=1}^K \frac{\hat{\alpha}_k}{\sum_{i=1}^K \hat{\alpha}_i} = 1, \quad \sum_{k=1}^K \langle s_{*,k} \rangle = 1 \right) \end{aligned}$$

3行目の式を予測分布 $p(\boldsymbol{x}_*|\boldsymbol{X}, \boldsymbol{S})$ で割ることで正規化できる。

実行環境

- Python 3.7
- Numpy 1.17
- Scipy 1.3
- Matplotlib 3.1
- Scikit-Learn 0.21

実行結果

上がVIGMM、下が通常のGMM。観測データが単純なので左図の予測分布にはあまり違いはない。右図のクラス予測はクラスごとの予測値をRGB値にマッピングしている。VIGMMは s_* がカテゴリカル分布に従う確率変数となるので期待値を求めることでクラス境界の曖昧さを表現できる。

