第11章 ベイズ推定法における モデル選択

11.3 ベイズ情報量規準
11.4 変分ベイズ法

06T4029T 斉藤優

ベイズ情報量規準(1)

- ラプラス近似の欠点
  - ヘッセ行列Hの計算が大変

$$\log p(\chi;\beta) \approx \sum_{i=1}^{n} \log q(x_i|\widehat{\theta}_{MAP}) + \log p(\widehat{\theta}_{MAP};\beta) + \frac{t}{2}\log(2\pi) - \frac{1}{2}\log(\det(-H))$$

- ラプラス近似の更なる近似

1. 第2項目と第3項目

訓練標本数nが十分大きいと仮定

$$\Downarrow$$

第2項目と第3項目の大きさはnに比例しないので無視

ベイズ情報量規準(2)

- ラプラス近似の更なる近似(続き)

  2. 第4項目
     大数の法則よりヘッセ行列Hの各要素の$\frac{1}{n}$倍は

$$
\begin{aligned}
\frac{1}{n}H_{i,j} &= \frac{\partial^2}{\partial\theta^{(i)}\partial\theta^{(j)}}\left(\frac{1}{n}\sum_{i=1}^{n}\log q(x_i|\theta)+\frac{1}{n}\log p(\theta;\beta)\right)\Bigg|_{\theta=\widehat{\theta}_{MAP}} \\
&\xrightarrow{p} \frac{\partial^2}{\partial\theta^{(i)}\partial\theta^{(j)}}\left(E_x[\log q(x|\theta)]\right)\Bigg|_{\theta=\widehat{\theta}_{MAP}} =: \widetilde{H}_{i,j}
\end{aligned}
$$

     $\widetilde{H}$は$t\times t$行列なので

$$
\det(n\widetilde{H}) = n^t\det(\widetilde{H})
$$

     よって第4項目は以下のように近似

$$
\frac{1}{2}\log(\det(-H)) \xrightarrow{p} \frac{t}{2}\log n + \underbrace{\frac{1}{2}\log\left(\det(-\widetilde{H})\right)}_{\text{n に依存しないので無視}}
$$

ベイズ情報量規準(3)

- ラプラス近似の更なる近似(続き)

3. 第1項目
最大事後推定量$\widehat{\theta}_{MAP}$と最尤推定量$\widehat{\theta}_{ML}$の差の大きさは高々n^{-1}に比例するので以下のように近似

$$\sum_{i=1}^{n} \log q(x_i|\widehat{\theta}_{MAP}) \approx \sum_{i=1}^{n} \log q(x_i|\widehat{\theta}_{ML})$$

4. これらをまとめると

$$\log p(\chi;\beta) \approx \sum_{i=1}^{n} \log q(x_i|\widehat{\theta}_{ML}) - \frac{t}{2}\log n$$

ベイズ情報量規準 (4)

- ベイズ情報量規準 (BIC)

$$BIC := -\sum_{i=1}^{n} \log q(x_i|\widehat{\theta}_{ML}) + \frac{t}{2}\log n$$

  – 事前確率の設定に用いることはできない

  – AIC と形が似ている

$$AIC := -\sum_{i=1}^{n} \log q(x_i;\widehat{\theta}_{ML}) + t$$

$$\Downarrow$$

どちらが優れているかは断定できない

変分ベイズ法(1)

- 周辺尤度がガウス関数に近くないとき

$\Longrightarrow$ ラプラス近似では精度が良くない

$\Longrightarrow$ 変分近似法が有効

変分近似法とは ⋯

あらかじめ用意しておいた計算しやすい確率密度関数集合の中から最も良い近似を求める手法

$\Downarrow$

この手法を利用して周辺尤度を近似

変分ベイズ法(2)

- 近似の求め方

  1. ある潜在変数ηを導入した場合の周辺尤度

  $$p(\chi;\beta) = \iint p(\chi,\eta,\theta;\beta)d\eta d\theta$$

  2. $p'(\eta)$と$p''(\theta)$という2つのモデルを導入して対数周辺尤度の下界を取得

  $$\begin{aligned}\log p(\chi;\beta) &= \log \iint p(\chi,\eta,\theta;\beta)d\eta d\theta \\ &= \log \iint p'(\eta)p''(\theta)\frac{p(\chi,\eta,\theta;\beta)}{p'(\eta)p''(\theta)}d\eta d\theta \\ &\geq \iint p'(\eta)p''(\theta)\log\frac{p(\chi,\eta,\theta;\beta)}{p'(\eta)p''(\theta)}d\eta d\theta =: F(p',p'')\end{aligned}$$

変分ベイズ法(3)

- 近似の求め方(続き)

2. (続き)
   $F(p', p'')$ を最大にするように $p'(\eta)$ と $p''(\theta)$ を決定
   $\Longrightarrow$ 対数周辺尤度の近似を取得
   $\Longrightarrow$ 関数に関する最大化問題は変分問題という

3. 下界 $F(p', p'')$ をゼロとおいた式

$$\frac{\partial}{\partial p'} F(p', p'') = 0, \quad \frac{\partial}{\partial p''} F(p', p'') = 0$$

変分ベイズ法(4)

- 近似の求め方(続き)

4. 3の式を解くと

$$p'(\eta) \propto \exp\left(\int p''(\theta)\log p(\chi,\eta|\theta;\beta)d\theta\right)$$

$$p''(\theta) \propto p(\theta)\exp\left(\int p'(\eta)\log p(\chi,\eta|\theta;\beta)d\eta\right)$$

この式を交互に計算して変分問題の解を求める

$$\Downarrow$$

変分ベイズEMアルゴリズム

p'を求めることはVB-Eステップ

p''を求めることはVB-Mステップ

変分ベイズ法(5)

- EM アルゴリズムにおける変分問題
  EM アルゴリズムも変分問題として解釈可能

1. 尤度の下界を取得

$$
\begin{aligned}
\log p(\chi|\theta) &= \log \int p(\chi, \eta|\theta) d\eta \\
&= \log \int p'(\eta|\chi, \theta') \frac{p(\chi, \eta|\theta)}{p'(\eta|\chi, \theta')} d\eta \\
&\geq \int p'(\eta|\chi, \theta') \log \frac{p(\chi, \eta|\theta)}{p'(\eta|\chi, \theta')} d\eta =: b(p', \theta)
\end{aligned}
$$

2. 下界をゼロとおいた式

$$
\frac{\partial}{\partial p'} b(p', \theta) = 0
$$

変分ベイズ法(6)

- EMアルゴリズムにおける変分問題(続き)

  3. 2の式を解く(Eステップ)

$$p'(\eta|\chi,\theta') = p(\eta|\chi,\theta)$$

  4. 以下の式を満たすθ'を決める(Mステップ)

$$\left.\frac{\partial}{\partial\theta} b(p',\theta)\right|_{\theta=\theta'} = 0_t$$

- VB-EステップとEステップの対応関係
  ディラックのデルタ関数

$$\int_{-\infty}^{\infty} g(\tau)\delta(\kappa-\tau)d\tau = g(\kappa)$$

変分ベイズ法(6)

- VB-E ステップと E ステップの対応関係(続き)
  デルタ関数は正規分布の確率密度関数の標準偏差がゼロの極限として表現可

$$\delta(\tau) = \lim_{\sigma \to 0} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\tau^2}{2\sigma^2} \right) \right)$$

  このデルタ関数を用いて $p''(\theta) = \delta(\theta)$ とおくと VB-E ステップは

$$p'(\eta) \propto p(\chi, \eta | \theta)$$

変分ベイズ法(7)

- VB-Eステップと Eステップの対応関係(続き)
  以下の式が成り立つので

$$p(\chi, \eta|\theta) = p(\eta|\chi, \theta)p(\chi|\theta) \propto p(\eta|\chi, \theta)$$

Eステップは

$$p'(\eta) \propto p(\eta|\chi, \theta)$$

よって, VB-EステップはEステップと一致する