

# 第9章 ベイズ推定法

9.3 最大事後確率推定法

9.4 共役事前分布

06T4029T 齊藤優

# 最大事後確率推定法(1)

- 最大事後確率推定法

- パラメータの事後確率を最大にするパラメータ値だけで確率密度関数を近似

$$\hat{p}_{MAP} := q(x|\hat{\theta}_{MAP}), \quad \hat{\theta}_{MAP} := \operatorname{argmax}_{\theta} p(\theta|\chi)$$

- 最大事後確率推定量

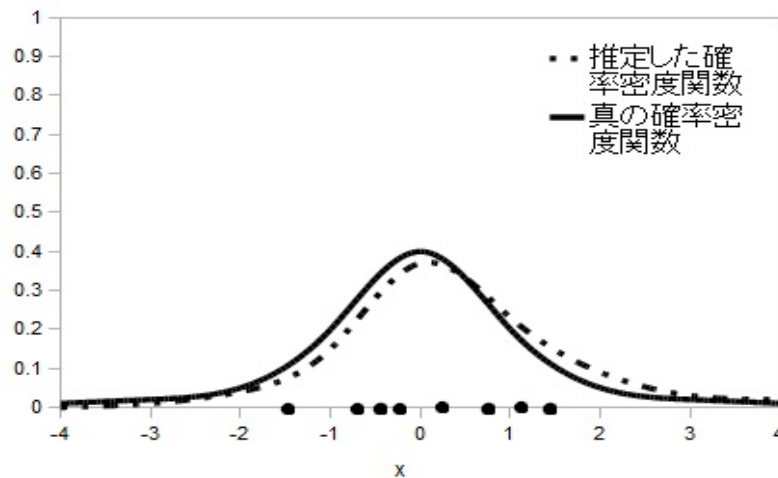
- 対数をとった事後確率を最大にするパラメータ値

$$\begin{aligned} \hat{\theta}_{MAP} &= \operatorname{argmax}_{\theta} \log p(\theta|\chi) \\ &= \operatorname{argmax}_{\theta} \left( \sum_{i=1}^n \log q(x_i|\theta) + \log p(\theta) \right) \end{aligned} \quad (1)$$

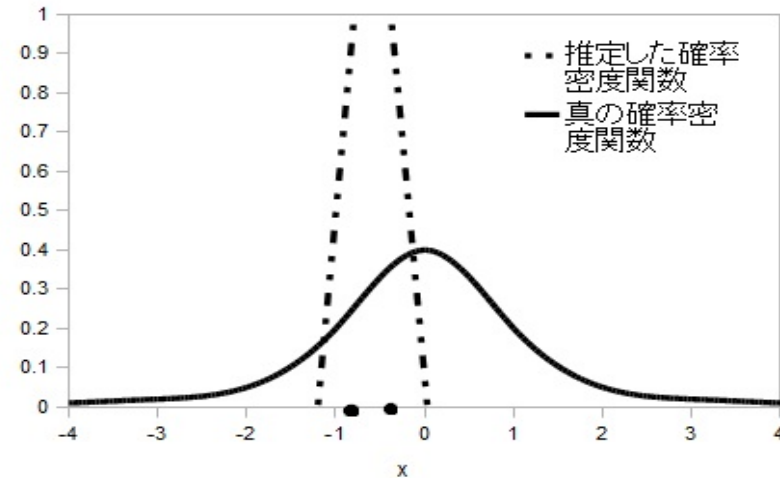
# 最大事後確率推定法(2)

- 最尤推定法の弱点

- 訓練標本数がモデルの複雑さと比べて相対的に少ない場合
    - ⇒ 尤度の値が非常に大きくなる
    - ⇒ 訓練標本に過剰に適応した確率密度関数が得られる (図1(b))
- このような現象は過適合とよばれる



(a)  $n=10$  のとき



(b)  $n=2$  のとき

図1. ガウス最尤推定の例

# 最大事後確率推定法(3)

- 罰則付き最尤推定法

- 最尤推定法は訓練標本数が少ない場合に尤度が大きくなり過ぎることに問題
  - ⇒ 過適合を抑制するため罰則が必要
  - ⇒ 式(1)の  $\log p(\theta)$  が罰則に対応
- 最大事後確率推定法は罰則付き最尤推定法
  - ⇒ 最大事後確率推定法は, 対数尤度だけでなく対数事前確率も合わせて大きくしなければならない
  - ⇒ うまく事前確率を選ぶことで過適合を効果的に抑制
- 罰則のことを正則化とよぶこともある

# 最大事後確率推定法(4)

- 最大事後確率推定量の例

- 分散共分散行列が単位行列  $I_d$  で固定のガウスモデル

$$q(x|\mu) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{(x-\mu)^\top(x-\mu)}{2}\right)$$

このモデルに対するガウス事前確率は

$$p(\mu; \gamma) = \frac{1}{(2\pi\gamma^2)^{\frac{d}{2}}} \exp\left(-\frac{\mu^\top\mu}{2\gamma^2}\right)$$

これは原点に近い  $\mu$  ほど事前確率が大きくなる

# 最大事後確率推定法 (5)

- 最大事後確率推定量の例 (続き)

- このときの最大事後確率推定量  $\hat{\mu}_{MAP}$

$$\hat{\mu}_{MAP} = \frac{1}{n + \gamma^{-2}} \sum_{i=1}^n x_i$$

最尤度推定量  $\hat{\mu}_{ML}$  は

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i$$

$\hat{\mu}_{MAP}$ ,  $\hat{\mu}_{ML}$ ,  $x_i$  の  $j$  番目の要素を  $\hat{\mu}_{MAP}^{(j)}$ ,  $\hat{\mu}_{ML}^{(j)}$ ,  $x_i^{(j)}$  で表すと,  $\gamma > 0$  より

$$\left| \hat{\mu}_{MAP}^{(j)} \right| = \frac{1}{n + \gamma^{-2}} \left| \sum_{i=1}^n x_i^{(j)} \right| < \frac{1}{n} \left| \sum_{i=1}^n x_i^{(j)} \right| = \left| \hat{\mu}_{ML}^{(j)} \right|$$

# 最大事後確率推定法 (6)

- 最大事後確率推定量の例 (続き)

- 上記のガウス事前確率を用いた最大事後確率推定量は, 最尤推定量よりも原点に近い

$$\|\hat{\mu}_{MAP}\| < \|\hat{\mu}_{ML}\|$$

⇒ 最大事後確率推定量は, 最尤推定量を事前確率が高い方に補正

- パラメータの事後確率の期待値を用いて確立密度関数を推定する方法

$$\hat{p}(x) := q(x|\hat{\theta}), \quad \hat{\theta} := \int \theta p(\theta|\chi) d\theta$$

# 共役事前分布 (1)

- 最大事後確率推定法と利点と欠点
  - 利点  $\implies$  計算が簡単であり過適合の抑制効果もある
  - 欠点  $\implies$  ベイズ推定法の性質が失われる
- 共役事前分布
  - 事前確率  $p(\theta)$  と事後確率  $p(\theta|\chi)$  が同じ種類の分布になるような確率分布
- ベイズ推定法の解の計算
  - 事後確率  $p(\theta|\chi)$  が閉じた形で求まっていると都合がよい
    - $\implies$  事前確率  $p(\theta)$  を共役事前分布に設定
    - $\implies$  事後確率の確率分布のパラメータを推定

# 共役事前分布 (2)

- 共役事前分布の具体例

- 期待値が0で分散 $\sigma^2$ の1次元ガウスモデル ( $\tau = \sigma^{-2}$ )

$$q(x|\tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau x^2}{2}\right)$$

ガンマ分布を事前確率に用いると

$$p(\tau; \alpha, \beta) = \frac{\tau^{\alpha-1} \beta^\alpha e^{-\beta\tau}}{\Gamma(\alpha)}$$

$\Gamma(\alpha)$  はガンマ関数で、正の整数  $k$  に対し以下を満たす

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \Gamma(k) = (k-1)!$$

# 共役事前分布 (3)

- 共役事前分布の具体例 (続き)

- モデルに対する尤度に対し, 事前確率を用いると, 事後確率は再びガンマ分布となる

$$p(\tau|\chi) = \frac{\tau^{\tilde{\alpha}-1} \tilde{\beta}^{\tilde{\alpha}} e^{-\tilde{\beta}\tau}}{\Gamma(\tilde{\alpha})}$$

ガンマ分布のパラメータは以下のように更新される

$$\tilde{\alpha} := \alpha + \frac{n}{2}, \quad \tilde{\beta} := \beta + \frac{\sum_{i=1}^n x_i^2}{2}$$

- 結論

- 共役事前分布を用いれば, 事後分布が閉じた形で計算でき都合が良い
- 共役でない事前分布を使いたい場合は, 近似的に事後分布を求めることになる