パターン認識と機械学習
2.3.6 ガウス分布に対するベイズ推論

茂木 哲矢

2008 年 6 月 27 日

目次

目的

- パラメータ上の事前分布を導入して，ベイズ主義的な扱いを導く．
- このファイルを作った感じでは，さまざまな場合における共役事前分布の紹介をしているような印象を受けた．

1 変数のガウス確率変数 x のベイズ主義 (1/3)

問題

分散 σ^2 は既知，N 個の観測値集合 $\boldsymbol{X} = \{x_1, \ldots, x_N\}$ から平均 μ を推定する．

尤度関数 $p(\boldsymbol{X}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \frac{1}{(2\pi\sigma)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$

事前分布 $p(\mu) = N(\mu|\mu_0, {\sigma_0}^2)$

事後分布 $p(\mu|\boldsymbol{X}) = N(\mu|\mu_N, {\sigma_N}^2)$

ただし，$\mu_N = \frac{\sigma^2}{N{\sigma_0}^2+\sigma^2}\mu_0 + \frac{N{\sigma_0}^2}{N{\sigma_0}^2+\sigma^2}\mu_{ML}$**，**

$\frac{1}{{\sigma_N}^2} = \frac{1}{{\sigma_0}^2} + \frac{N}{\sigma^2}$ **，** $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$

事後分布の平均と分散

平均 $\mu_N = \begin{cases} \frac{\sigma^2}{\sigma^2}\mu_0 + \frac{0}{\sigma^2}\mu_{ML} = \mu_0 & (N=0) \\ \frac{\sigma^2}{\infty} + \frac{{\sigma_0}^2}{{\sigma_0}^2+\sigma^2/\infty}\mu_{ML} = \mu_{ML} & (N=\infty) \end{cases}$

分散 $\frac{1}{{\sigma_N}^2} = \begin{cases} \frac{1}{{\sigma_0}^2} & (N=0) \\ \frac{1}{{\sigma_0}^2} + \frac{\infty}{\sigma^2} & (N \to \infty) \end{cases}$

有限の N の場合，分散が無限 $({\sigma_0}^2 \to \infty)$ の事前分布で，

平均 $\mu_N = \frac{\sigma^2}{\infty+\sigma^2}\mu_0 + \frac{N}{N+\sigma^2/\infty}\mu_{ML} = \mu_{ML}$

分散 $\frac{1}{{\sigma_N}^2} = \frac{1}{\infty} + \frac{N}{\sigma^2} = \frac{N}{\sigma^2}$

ベイズ推論への逐次推定の導入

逐次推定

N 個のデータ点に対する問題を N 番目のデータ点 x_N の影響と $N-1$ 個のデータ点の問題の組み合わせと考える．

$$p(\mu|\boldsymbol{X}) \propto \left[p(\mu) \prod_{n=1}^{N-1} p(x_n|\mu) \right] p(x_N|\mu)$$

利点

- 非常に汎用的．
- データが独立同分布に従うとすると，どんな問題にも適用できる．

1 変数のガウス確率変数 x のベイズ主義 (2/3)

問題

平均を既知として，分散を推定する．

精度 $\lambda = 1/\sigma^2$ **として，**

尤度関数 $p(\boldsymbol{X}|\lambda) = \prod_{n=1}^{N} N(x_n|\mu, \lambda^{-1}) \propto \lambda^{N/2} \exp\left\{-\frac{\lambda}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$

共役事前分布 **ガンマ分布**

$$Gam(\lambda|a_0, b_0) = \frac{1}{\Gamma(a_0)}{b_0}^{a_0}\lambda^{a_0-1}\exp(-b_0\lambda)$$

事後分布 $p(\lambda|\boldsymbol{X}) \propto \lambda^{a_0-1}\lambda^{N/2}\exp\left\{-b_0\lambda - \frac{\lambda}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\} = Gam(\lambda|a_N, b_N)$

ただし， $a_N = a_0 + \frac{N}{2}$**，**

$$b_N = b_0 + \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2 = b_0 + \frac{N}{2}{\sigma_{ML}}^2$$

ガンマ分布のパラメータ

N 個のデータ点を観測すると，

- $a_N = a_0 + \frac{N}{2}$

  $\rightarrow a$ を $N/2$ だけ増やす．

  $\rightarrow a_0$ は，$2a_0$ 個の「有効な」観測値があった．

- $b_N = b_0 + \frac{N}{2}{\sigma_{ML}}^2$

  $\rightarrow b$ を $N{\sigma_{ML}}^2/2$ だけ増やす．

  $\rightarrow b_0$ は分散が $2b_0/(2a_0) = b_0/a_0$ であるような，$2a_0$ 個の「有効な」観測値が事前にあった.

指数型分布族では一般的に，共役事前分布が有効な仮想データ点と解釈できる．

（分散そのもので考える場合の共役事前分布は逆ガンマ分布である.)

1変数のガウス確率変数 x のベイズ主義 (3/3)

問題

平均と分散の両方が未知の場合の共役事前分布を求めるために，尤度関数の μ と λ の依存関係を調べる．

尤度関数 $p(\boldsymbol{X}|\mu,\lambda) = \prod_{n=1}^{N}(\frac{\lambda}{2\pi})^{1/2}\exp\left\{\frac{\lambda}{2}(x_n-\mu)^2\right\} \propto$

$\left[\lambda^{1/2}\exp(-\frac{\lambda\mu^2}{2})\right]^N \exp\left\{\lambda\mu\sum_{n=1}^{N}x_n - \frac{\lambda}{2}\sum_{n=1}^{N}{x_n}^2\right\}$

事前分布 $p(\mu,\lambda) \propto \left[\lambda^{1/2}\exp(-\frac{\lambda\mu^2}{2})\right]^\beta \exp\{c\lambda\mu - d\lambda\} =$

$\exp\left\{-\frac{\beta\lambda}{2}(\mu-c/\beta)^2\right\}\lambda^{\beta/2}\exp\left\{-(d-\frac{c^2}{2\beta})\lambda\right\}$

(c,d,β **：定数)**

正規-ガンマ分布やガウス-ガンマ分布

$p(\mu,\lambda) = N(\mu|\mu_0,(\beta\lambda)^{-1})Gam(\lambda|a,b)$

($a = 1+\beta/2, b = d - c^2/2\beta$**)**

D 次元変数の多変量ガウス分布 $N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$ の場合

平均	精度	共役事前分布
未知	既知	ガウス分布
既知	未知	ウィシャート分布 $\mathcal{W}(\boldsymbol{\Lambda}\|\boldsymbol{W}, \nu) = B\|\boldsymbol{\Lambda}\|^{(\nu-D-1)/2} \exp(-\frac{1}{2}\mathbf{Tr}(\boldsymbol{W}^{-1}\boldsymbol{\Lambda}))$ (精度行列ではなく共分散行列を使うと，逆ウィシャート分布)
未知	未知	正規-ウィシャート分布・ガウス-ウィシャート分布 $p(\boldsymbol{\mu}, \boldsymbol{\Lambda}\|\boldsymbol{\mu_0}, \beta, \boldsymbol{W}, \nu) = N(\boldsymbol{\mu}\|\boldsymbol{\mu_0}, (\beta\boldsymbol{\Lambda})^{-1})\mathcal{W}(\boldsymbol{\Lambda}\|\boldsymbol{W}, \nu)$