パターン認識と機械学習

2.4.3 無情報事前分布

吉田拓夢

無情報事前分布

確率的推論の適用時

- 事前知識→事前分布
- 事前分布:ある変数の値→確率0
  事後分布:　　″　　　　　→確率0

分布の形状について知見が少ないときは

- 事後分布への影響が極力ない→無情報事前分布

定数の事前分布

- p(x|λ)において事前分布p(λ)=constとしたい

→離散:

λがK個の状態→事前分布は各々1/K

→連続:

2つの問題点

問題点

- λの定義域が有界でない→λ上の積分が発散
  →事前分布は正しく正規化できない

↑変則事前分布/不完全事前分布

- 変則事前分布でも正しく正規化できればOK

例：ガウス分布の平均パラメータ上での均一事前分布

問題点

- 非線形な変数変換をしたとき

$$p_\eta(\eta) = p_\lambda(\lambda)\left|\frac{\mathrm{d}\lambda}{\mathrm{d}\lambda}\right| = p_\lambda(\eta^2)2\eta \propto \eta$$

λ=η^2,pλ(λ)を定数としてもη上の密度は定数でない

無情報事前分布の例1

$$p(x|\mu) = f(x - \mu)$$

ガウス分布の平均μなど

- μ:位置パラメータ
- 平行移動不変性(並進不変性)を持つ

$\hat{x} = x + c$　とすると

$$p(\hat{x}|\hat{\mu}) = f(\hat{x} - \hat{\mu})$$

ただし　$\hat{\mu} = \mu + c$

無情報事前分布の例1

- 平行移動不変性を持つ事前分布

  →区分　$A \le \mu \le B$　に入る確率と区分

  $A - c \le \mu \le B - c$　に入る確率が等しい

$$\int_A^B p(\mu)\mathrm{d}\mu = \int_{A-c}^{B-c} p(\mu)\mathrm{d}\mu = \int_A^B p(\mu - c)\mathrm{d}\mu$$

任意のA,Bについて成立するので

$$p(\mu - c) = p(\mu)$$

p(μ)は定数となる

無情報事前分布の例2

$$p(x|\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$

ガウス分布の
標準偏差σなど

- σ:尺度パラメータ
- 尺度不変性を持つ

$\hat{x} = cx$ とすると

$$p(\hat{x}|\hat{\sigma}) = \frac{1}{\hat{\sigma}} f\left(\frac{\hat{x}}{\hat{\sigma}}\right)$$

ただし $\hat{\sigma} = c\sigma$

無情報事前分布の例2

- 尺度不変性を持つ事前分布

  →区分　$A \leq \sigma \leq B$　に入る確率と区分

  $A/c \leq \sigma \leq B/c$　に入る確率が等しい

$$\int_A^B p(\sigma)\mathrm{d}\sigma = \int_{A/c}^{B/c} p(\sigma)\mathrm{d}\sigma = \int_A^B p\left(\frac{1}{c}\sigma\right)\frac{1}{c}\mathrm{d}\sigma$$

任意のA,Bについて成立するので

$$p(\sigma) = p\left(\frac{1}{c}\sigma\right)\frac{1}{c}$$

無情報事前分布の例2

$p(\sigma) = p\left(\frac{1}{c}\sigma\right)\frac{1}{c}$ より、条件 $p(\sigma) \propto 1/\sigma$ を得る

- 変則事前分布である

  $0 \le \sigma \le \infty$ での分布の積分で発散

- 尺度パラメータの事前分布ではパラメータの対数の密度を考えると便利

  p(ln σ)=const