続・わかりやすいパターン認識
-教師なし学習入門-

第4章 パラメータ推定

河野 和平

パラメータ推定

- 線形識別関数

$$g_i = w_{i0} + \sum_{j=1}^{d} w_{ij} x_j \quad (i = 1,2,\ldots,c)$$

  - 重み係数 $w_{i0}, w_{i1}, \ldots, w_{id}$

パーセプトロンの学習規則
誤差逆伝播法

- ベイズ識別関数

$$g_i\left(x^{(n)}\right) = P(w_i) \cdot P\left(x^{(n)} \middle| w_i\right)$$

  - 確率関数 $P\left(x^{(n)} \middle| w_i\right)$，事前確率 $P(w_i)$

パラメータ推定

コイン投げの例

- コイン投げ
  - 表の出る確率θ
  - n 回投げた結果$x^{(n)} = x_1 x_2 \cdots x_n$
  - n 回のうち r 回が表

- 観測結果 $x^{(n)}$が得られる確率$P\left(x^{(n)};\theta\right)$

$$P\left(x^{(n)};\theta\right) = \prod_{t=1}^{n} P(x_t;\theta)$$

$$= \theta^r (1-\theta)^{n-r}$$

最尤推定

- $P\left(x^{(n)};\theta\right)$ を θ について最大化

$$\hat{\theta} = \arg\max_{\theta}\left\{P\left(x^{(n)};\theta\right)\right\}$$

$$\frac{d}{d\theta} log\, P\left(x^{(n)};\theta\right) = \frac{d}{d\theta}\left(rlog\theta + (n-r)log(1-\theta)\right)$$

$$= \frac{r}{\theta} - \frac{n-r}{1-\theta} = 0$$

$$\hat{\theta} = \frac{r}{n}$$

ベイズ推定1

- パラメータ θ を確率変数とみなす
- 観測結果 $x^{(n)}$ が得られたときの条件付き確率密度

$$p\left(\theta \middle| x^{(n)}\right) = \frac{P\left(x^{(n)} \middle| \theta\right)}{P\left(x^{(n)}\right)} \cdot p(\theta)$$

  – パラメータ θ は連続的な値を取る

$$P\left(x^{(n)}\right) = \int_0^1 P\left(x^{(n)} \middle| \theta\right) p(\theta)\, d\theta$$

$$\int_0^1 p(\theta)\, d\theta = 1$$

$$\int_0^1 p\left(\theta \middle| x^{(n)}\right) d\theta = 1$$

$$\sum_{x^{(n)}} P\left(x^{(n)}\right) = 1$$

ベイズ推定2

- 観測結果 $x^{(n)}$ が得られる確率 $P(x^{(n)}|\theta)$

$$P(x^{(n)}|\theta) = \prod_{t=1}^{n} P(x_t|\theta)$$

$$= \theta^r (1-\theta)^{n-r}$$

… 尤度

- 表の出る確率 θ についての情報がない
  - 一様分布を仮定

$$p(\theta) = 1$$

ベイズ推定3

$$P\left(x^{(n)}\right) = \int_0^1 \theta^r (1-\theta)^{n-r} \cdot 1\, d\theta$$

$$= B(r+1, n-r+1)$$

$$= \frac{r!\,(n-r)!}{(n+1)!}$$

$$= \frac{1}{(n+1) \cdot {}_n C_r}$$

$$p\left(\theta \middle| x^{(n)}\right) = (n+1) \cdot {}_n C_r\, \theta^r (1-\theta)^{n-r}$$

・・・ベータ分布

共役事前分布

- ベイズ更新
  - 事前分布：一様分布
  - 事後分布：ベータ分布 $Be(\alpha, \beta)$

⇩

  - 事前分布：ベータ分布 $Be(\alpha, \beta)$

$$p(\theta) = Be(\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$$

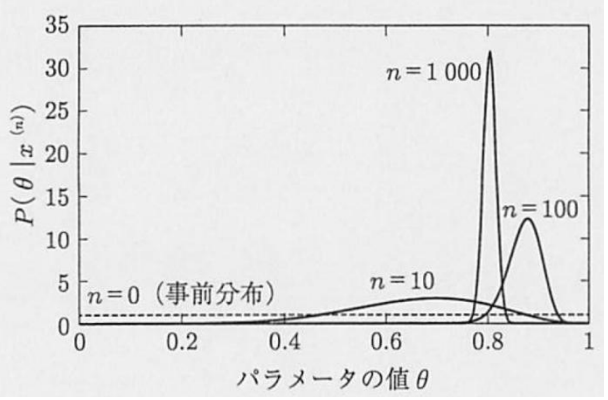
  - 事後分布：

$$p(\theta|x^{(n)}) = \frac{1}{Z_1}\theta^{\alpha+r-1}(1-\theta)^{\beta+n-r-1}$$

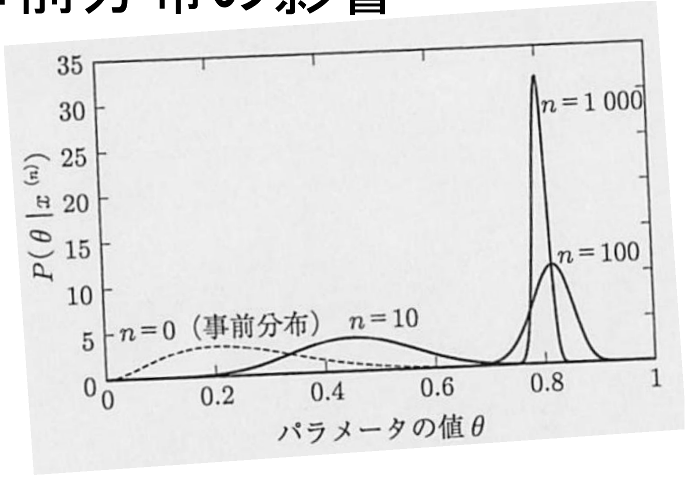
$$= Be(\alpha+r, \beta+n-r) \qquad \cdots \text{ベータ分布}$$

推定結果

- 観測回数の効果



- 事前分布の影響



ディリクレ分布

- ベータ分布の一般形

$$Dir(\alpha_1, \dots, \alpha_m) = \frac{\Gamma(n+m)}{\Gamma(n_1+1)\cdots\Gamma(n_m+1)}\theta_1^{n_1}\cdots\theta_m^{n_m}$$

- 多項分布のパラメータ θ のベイズ推定
  - 事前分布: $p(\theta) = Dir(\alpha_1, \dots, \alpha_m)$
  - 事後分布:

$$p(\theta | x^{(n)}) = \frac{1}{Z_3} \cdot \theta_1^{\alpha_1+n_1-1} \cdots \theta_m^{\alpha_m+n_m-1}$$

$$= Dir(\alpha_1 + n_1, \dots, \alpha_m + n_m)$$