P-5

ベイズ推定とギブズ推定
Bayes and Gibbs estimation

○岡田憲相 [1], 青柳美輝 [2]
Kensuke Okada, Miki Aoyagi

Abstract: In statistical learning theory, one of its main purposes is to measure the difference between the true density function and the estimated one for a given method of statistical estimation. In this paper, we show one of the main formulas in singular learning theory, for clarifying mathematical relations in the Bayes quartet, i.e., the means of Bayes generalization error B_g, Bayes training error B_t, Gibbs generalization error G_g and Gibbs training error G_t.

1. はじめに

情報科学の分野では, 計測して得られた多量のデータから元の情報源の確率分布を推測することを, 学習という. 学習の仕組みをまとめて体系化したものを学習理論といい, 手書き文字をはじめとする画像認識, 音声認識や遺伝子解析などに応用されている. これらの学習データは，非常に複雑な構造を持っているため, 学習モデルも十分複雑なものを考える必要がある．そのような学習モデルとして，階層型ニューラルネットワーク，混合正規分布，隠れマルコフモデルといったものがあり，特異モデルと呼ばれている．特異モデルにおいて, 学習により得られた確率密度関数と真の確率密度関数との差異を考えることが本稿の目的である. 特にベイズ, ギブズそれぞれの汎化誤差, 学習誤差の関係を述べていく.

2. ベイズカルテット

$x \in \mathbf{R}^N$，$w \in W \subset \mathbf{R}^d$ に対して，$q(x)$ と $p(x|w)$ は同じサポートを持つ $\mathbf{R}^N$ 上の確率密度関数とする．ここで $p(x|w)$ はパラメータ $w \in W \subset \mathbf{R}^d$ に依存するパラメータ付きの確率密度関数である. また, パラメータ集合 W は $\mathbf{R}^d$ でコンパクトとする. 学習モデル $p(x|w)$ とその事前分布 $\varphi(w)$ が与えられているものとし，真の分布を $q(x)dx$ とする．以下, 確率分布 $q(x)dx$ に従う確率変数 $X_1, \cdots, X_n$ は独立であるとし, 確率変数の集合 $D_n = \{X_1, \cdots, X_n\}$ に対して, 正規事後分布を,

$$p(w|D_n) = \frac{1}{Z_n}\varphi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta}$$

で定義する. ここで $\beta > 0$ は逆温度である. Z_n は正規化定数とよばれ, $Z_n = \int dw\varphi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta}$ と定義される.

定義 1 (ベイズカルテット) $E_w[\cdot]$ を $p(w|D_n)$ を用いた期待値とし, 以下の 4 つの誤差を定義する.

(1) ベイズ汎化誤差 $B_g = E_X\left[\log \frac{q(X)}{E_w[p(X|w)]}\right]$

(2) ベイズ学習誤差 $B_t = \frac{1}{n}\sum_{i=1}^{n} \log \frac{q(X_i)}{E_w[p(X_i|w)]}$

(3) ギブズ汎化誤差 $G_g = E_w\left[E_X\left[\log \frac{q(X)}{p(X|w)}\right]\right]$

(4) ギブズ学習誤差 $G_t = E_w\left[\frac{1}{n}\sum_{i=1}^{n} \log \frac{q(X_i)}{p(X_i|w)}\right]$

これら 4 つの誤差の組をベイズカルテットという.

定義 2 (基本条件) カルバック距離を,

$$K(w) = \int q(x)\log\frac{q(x)}{p(x|w)}dx$$

で定義する. ここで, $W_0 = \{w \in W; K(w) = 0\}$ は空集合でないとする. 次の条件を満たすとき, $q(x)$ と $p(x|w)$ は指数 s $(s \geq 2)$ で基本条件を満たすという.
(1) $f(x,w) = \log(q(x)/p(x|w))$ に対して,
[1-a] $\mathbf{C}^d$ の開集合 $W^{(C)}$ が存在して, $W \subset W^{(C)}$.
[1-b] $W^{(C)} \ni w \mapsto f(\cdot, w)$ は $L^s(q)$ 値の複素解析関数.
[1-c] $M(x) \equiv \sup_{w \in W^{(C)}} |f(x,w)|$ は $L^s(q)$ に含まれる.
(2) $Q(x) \equiv \sup_{K(w) \leq \epsilon} p(x|w)$ に対して次の積分が,
$\int M(x)^2 Q(x)dx < \infty$ となるような $\epsilon > 0$ が存在する.

定義 3 特異点解消定理 *(Hironaka Theorem[2])* を，カルバック距離 $K(w)$ に適用すれば，ある多様体 $\mathscr{M}$ 上から W への写像 π が存在して，$\mathscr{M}$ のそれぞれの適当な局所座標 $u = (u_1, u_2, \cdots, u_d)$ において $K(\pi(u)) = u_1^{2k_1}\cdots u_d^{2k_d}$，$|\pi'(u)| = u_1^{h_1}\cdots u_d^{h_d}$ と書ける．このとき，$L^s(q)$ 値をとる解析関数 $a(x,u)$ が存在して，$f(x,\pi(u)) = a(x,u)u_1^{k_1}\cdots u_d^{k_d}$ となる.

解析多様体 $\mathscr{M}$ 上の正規分布 ξ を $E[\xi(u)] = 0$, $E[\xi(u)\xi(v)] = E_X[a(X,u)a(X,v)] - u_1^{k_1}\cdots u_d^{k_d}v_1^{k'_1}\cdots v_d^{k'_d}$ となるようにとる.

λ を学習係数すなわち $\min_u \min_i \frac{h_i+1}{2k_i}$ とし，その最大の位数を r とする.
$\frac{h_1+1}{2k_1} = \frac{h_2+1}{2k_2} = \cdots = \frac{h_r+1}{2k_r} = \lambda$,
$\frac{h'_j+1}{2k'_j} > \lambda \quad (j > r)$ が成立する局所座標 u^* におい

1: 日大理工・院 (前)・数学　2: 日大理工・教員・数学

て，$y' = (u_1, \cdots, u_r), y = (u_{r+1}, \cdots, u_d)$ とおく．また，$a_0(X, y) = a(X, 0, y), \xi_0(y) = \xi(0, y)$ とする．

このとき

$$\begin{aligned} B_g^*(\xi) &\equiv \frac{1}{2} E_X[E_{y,t}[a_0(X,y)t^{1/2}|\xi]^2] \\ B_t^*(\xi) &\equiv G_t^*(\xi) - G_g^*(\xi) + B_g^*(\xi) \\ G_g^*(\xi) &\equiv E_{y,t}[t|\xi] \\ G_t^*(\xi) &\equiv E_{y,t}[t - t^{1/2}\xi_0(y)|\xi] \end{aligned}$$

と定義する. ここで $E_{y,t}[f(y,t)|\xi]$ は,

$$Z_0(y,t,\xi) = y^\mu t^{\lambda-1} e^{-\beta t + \beta\sqrt{t}\xi_0(y)} \phi_0^*(y)$$

とおいたとき,

$$E_{y,t}[f(y,t)|\xi] = \frac{\sum_{u^*} \int_0^\infty dt \int dy f(y,t) Z_0(y,t,\xi)}{\sum_{u^*} \int_0^\infty dt \int dy Z_0(y,t,\xi)}$$

と定義される.

$$\phi_0^*(y) \equiv \sigma_{\alpha,0}(y)\phi_0(y), \phi_0(y) = \varphi(\pi(0,y)) \in C^\infty$$

であり, C^∞ の関数集合 $\{\sigma_{\alpha,0}(y)\}$ は $\mathscr{M}$ の 1 の分解である.

定理 4 $s = 6$ で基本条件が成立するとする.

(1) $n \to \infty$ のとき, 以下の法則収束が成り立つ.

$$nB_g \to B_g^*,\ nB_t \to B_t^*,\ nG_g \to G_g^*,\ nG_t \to G_t^*$$

(2) $n \to \infty$ のとき, 以下の確率収束が成り立つ.

$$n(B_g - B_t - G_g + G_t) \to 0$$

(3) ベイズカルテットの期待値は以下のように収束する.

$$E[nB_g] \to E[B_g^*], \quad E[nB_t] \to E[B_t^*],$$
$$E[nG_g] \to E[G_g^*], \quad E[nG_t] \to E[G_t^*]$$

定理 5 (統計的推測の状態方程式) $s = 6$ で基本条件が成立するとき, 次の等式が成り立つ.

$$\begin{aligned} E[B_g^*] - E[B_t^*] &= 2\beta(E[G_t^*] - E[B_t^*]), \\ E[G_g^*] - E[G_t^*] &= 2\beta(E[G_t^*] - E[B_t^*]). \end{aligned}$$

定理 6 $s = 6$ で基本条件が成り立つとする. $n \to \infty$ のとき確率収束,

$$nG_g + nG_t - \frac{2\lambda}{\beta} \to 0$$

が成り立つ. さらに,

$$E[G_g^*] + E[G_t^*] = \frac{2\lambda}{\beta}$$

が成り立つ.

定義 7 対数尤度関数の分散 V は,

$$V = \sum_{i=1}^n \left\{ E_w[(\log p(X_i|w))^2] - (E_w[\log p(X_i|w)])^2 \right\}$$

と定義される.

対数密度関数 $f(x,w) = \log(q(x)/p(x|w))$ を用いれば, 次のように書き表すことが出来る.

$$V = \sum_{i=1}^n \left\{ E_w[f(X_i,w)^2] - E_w[f(X_i,w)]^2 \right\}$$

定理 8 以下の確率収束が成り立つ.

$$\begin{aligned} V - 2(nG_t - nB_t) &\to 0 \\ V - 2(nG_g - nB_g) &\to 0 \end{aligned}$$

ある定数 $\nu = \nu(\beta) > 0$ が存在して,

$$\lim_{n\to\infty} E[V] = \frac{2\nu(\beta)}{\beta}$$

また,

$$\begin{aligned} E[B_g^*] &= \frac{\lambda}{\beta} + \left(1 - \frac{1}{\beta}\right)\nu(\beta) \\ E[B_t^*] &= \frac{\lambda}{\beta} - \left(1 + \frac{1}{\beta}\right)\nu(\beta) \\ E[G_g^*] &= \frac{\lambda}{\beta} + \nu(\beta) \\ E[G_t^*] &= \frac{\lambda}{\beta} - \nu(\beta) \end{aligned}$$

定数 $\nu(\beta) > 0$ を特異ゆらぎといい, 次の不等式を満たす.

$$0 \le \nu(\beta) \le \frac{E_\xi[\|\xi\|^2]}{8} + \frac{E_\xi[\|\xi\|^2]^{1/2}}{8}\sqrt{E_\xi[\|\xi\|^2] + 16\lambda/\beta}$$

ここで $\nu(\beta)$ は,

$$\nu(\beta) = \frac{1}{2} E_\xi\left[E_{y,t}[t^{1/2}\xi_0(y)|\xi]\right]$$

と定義され, $\|\xi\|$ は不規則過程 $\xi(u)$ の最大値を表す.

3. 参考文献

[1] S. Watanabe: “ Algebraic Geometry and Statistical Learning Theory ”, Cambridge University Press, New York, USA, vol.25, 2009.

[2] H. Hironaka: “ Resolution of Singularities of an algebraic variety over a field of characteristic zero ”, Annals of Math., vol.79, pp.109-326, 1964.

[3] 舟木直久, 確率論, 朝倉書店, p.261, 2004.