P-11

学習理論における特異性と特異点解消について
Singularities in learning theory and desingularization

○高村京輔 [1], 青柳美輝 [2]
*Kyosuke Takamura[1], Miki Aoyagi[2]

Abstract: In recent studies, statistical learning theory is analyzed by methods in algebraic geometry, because singular learning models have a singular Fisher metric that is not always approximated by any quadratic form. In this paper, we discuss the behaviors of empirical processes. Such processes are empirical Kullback distances, which relate to training errors. We show that the desingularization has an important role in analyzing such behaviors. After resolution, we have a common standard form of the empirical processes and its convergence form in law by using the central limit theorem.

1. 導入

画像音声認識，確率推論，遺伝子解析，時系列予測，データマイニングなどの学習システムの目的は，学習データから，学習モデルを用いて，学習データを発している真の分布を推測することである．このような学習データは，正規分布に従っているような単純なものではなく，非常に複雑な構造をもっている．したがって，学習モデルは十分複雑なものを考えなくてはならない．現在，そのようなデータを扱える学習モデルとして，神経回路網，混合正規分布，縮小ランクモデルなどの階層構造，内部構造を持つものが応用され，正規分布のような“よい性質”を持つ統計的正則モデルと区別するため，特異モデルと呼ばれている．近年，学習効率を表す学習係数の研究において，代数幾何の特異点解消（広中定理）との関係が明らかになり，急速に多くの理論の研究が始まった．この論文では，特異点解消がどのように学習係数に利用されるかを見る．

2. 特異点解消

定義 1 (特異点解消) $f(x)$ を $\mathbb{R}^d$ における原点 0 の近傍で定義された解析関数とし，$Z=f^{-1}(0)$ とする．原点 0 の近傍 U と，特異点を持たないある多様体 Y と，Y から U への固有な解析関数 π が，次の (1)，(2)，(3) を満たすとき，特異点解消という．

(1) $E=\pi^{-1}(Z)$ に対して $\pi : Y\setminus E \to U\setminus Z$ は同型を与える．

(2) 任意の $y\in Y$ について十分小さな近傍 V と適当な局所座標 $(y_1,\cdots,y_d)$ が存在して，

$$f(\pi(y)) = y_1^{s_1}y_2^{s_2}\cdots y_d^{s_d}$$

が成り立つ．ここで $s_1,\cdots,s_d$ は非負の整数．

(3) 写像 π のヤコビアン J_π は非負の整数 $m_1,\cdots,m_d$ が存在して，

$$J_\pi(y_1,\cdots,y_d) = y_1^{m_1}\cdots y_d^{m_d}\tilde{J}_\pi(y_1,\cdots,y_d)$$

と書ける．ここで，$\tilde{J}_\pi(0,\cdots,0)\neq 0$ である．

また，(1) の代わりに，次の (1′) と (2)，(3) が成り立つとき，π を広義の特異点解消であるという．

(1′) U の解析的真部分集合 N が存在し，$\pi : Y\setminus\pi^{-1}(N)\to U\setminus N$ は同型.

広中の定理は次のように述べられる．

定理 2 (広中 [2]) すべての解析関数 f について特異点解消が存在する．

定義 3 (舟木 [1]) $(\Omega,\mathcal{B})$ を可測空間とする．$\{P_n\}$ と P を確率分布の列および確率分布とする．$\{P_n\}$ が P に法則収束 (弱収束) するとは，任意の有界連続関数 $f:\Omega\to\mathbb{R}^1$ に対して，

$$\int f(x)P(dx) = \lim_{n\to\infty}\int f(x)P_n(dx)$$

が成り立つ時である．また，$(\Omega,\mathcal{B})$ に値を取る確率変数の列 $\{X_n\}$ が確率変数 X に法則収束 (弱収束) するとは，X_n の確率分布 P_{X_n} が X の確率分布 P_X に法則収束するときである．

定義 4 (舟木 [1]) $K\subset\mathbb{R}^d$ をコンパクト集合，$C(K)$ を K から $\mathbb{C}$ への連続関数の全体とする．$f_1,f_2\in C(K)$ に対して $d(f_1,f_2)=\max\{|f_1(x)-f_2(x)| : x\in K\}$ によって距離を定義する．$\mathcal{B}$ を $C(K)$ のボレル集合とする．

(1) 確率測度 μ がタイトであるとは，任意の $\epsilon>0$ に対して，あるコンパクト集合 $C\subset C(K)$ が存在して，$\mu(C)>1-\epsilon$ が成り立つ．

(2) 確率測度の列 $\{\mu_n\}$ が一様にタイトであるとは，任意の $\epsilon>0$ に対して，あるコンパクト集合 $C\subset C(K)$ が存在して，$\inf_n\mu_n(C)>1-\epsilon$ が成り立つ．

1: 日大理工・院 (前)・数学　2: 日大理工・教員・数学

(3) $C(K)$ に値を取る，確率変数 X がタイトであるとは，X の確率測度が，タイトであるときである．確率変数 X_n が一様にタイトであるとは，$\{X_n\}$ の確率測度の列 $\{\mu_n\}$ が一様にタイトであるときである．

定義 5 $s \geqq 1$ を実数, $W \subset \mathbb{R}^d$ を開集合とする．$q(x)$ を $\mathbb{R}^N$ 上の確率分布関数とする．関数 $f: \mathbb{R}^N \times W \ni (x,w) \mapsto f(x,w) \in \mathbb{R}$ が，w を固定したとき，x の解析関数であり，$\int |f(x)|^s q(x)dx < \infty$ であるとき，f を W 上の $L^s(q)$ 値解析関数という．

定理 6 (Watanabe[3]) s を正の偶数とする．開集合 $W \subset \mathbb{R}^d$ に対して $f: \mathbb{R}^N \times W \ni (x,w) \mapsto f(x,w) \in \mathbb{R}$ を W 上の $L^s(q)$ 値解析関数とする．$K \subset W$ をコンパクト集合とする．このとき，K 上の連続関数

$$\psi_n(w) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i,w) - E_X[f(X,w)]\}$$

は，次を満たす.

(1) $\{\psi_n(w)\}$ は一様にタイトである.

(2) $\{\psi_n(w)\}$ は, $C(K)$ 値の関数 $\psi(w)$ に法則収束.

(3) $E[\psi_n(w)] = 0$, $E[\psi_n(w)\psi_n(w')]$
$= E_X[(f(X,w) - m(w))\,(f(X,w') - m(w'))]$.
ここで，$m(w) = E_X[f(X,w)]$.

3. 学習理論

$x \in \mathbf{R}^N$，$w \in W \subset \mathbf{R}^d$ に対して，学習モデル $p(x|w)$ とその事前分布 $\psi(w)$ が与えられているものとし，真の分布を $q(x)$ とする．$q(x)$ に従う独立なサンプルを $D_n = \{X_1, X_2, ..., X_n\}$ とすると，事後確率 $p(w|D_n)$ は

$$p(w|D_n) = \frac{1}{Z_n}\psi(w)\prod_{i=1}^{L} p(X_i|w)$$

で与えられる（Z_n は正規化定数）．これよりベイズ推測は

$$p(x|D_n) = \int p(x|w)p(w|D_n)dw$$

と定義される．カルバック距離を

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

経験カルバック距離を

$$K_n(w) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{q(X_i)}{p(X_i|w)}$$

とおく. $E[K_n(w)] = K(w)$ が成り立つ.

特異点解消定理 (Hironaka Theorem) を，カルバック距離 $K(w)$ に適用すれば，任意の $w \in K^{-1}(0) \cap W$ に対して，ある近傍 U_w と，ある多様体 Y_w が存在して，Y_w の局所座標 $u = (u_1, u_2, \cdots, u_d)$ において，$K(\pi(u)) = u_1^{2k_1} \cdots u_d^{2k_d}$ と書ける.

$$f(x,w) = \log \frac{q(x)}{p(x|w)}$$

とおく.

定理 7 (Watanabe[3]) $s = 2$ に対して，次の条件が成り立つと仮定する．ある開集合 $W^{(C)} \supset W$ が存在して，$W^{(C)}$ 上 $f(x,w)$ は $L^s(q)$ 値解析関数．$M(x) = \sup_{w \in W^{(C)}} |f(x,w)|$ は $L^s(q)$ 関数．また，ある $\epsilon > 0$ が存在して，

$$\int M(x)^2 (\sup_{K(w)<\epsilon} p(x|w)) dx < \infty.$$

このとき，以下が成り立つ.

(1) $L^s(q)$ 値をとる解析関数 $a(x,u)$ が存在して，

$$f(x,\pi(u)) = a(x,u) u_1^{k_1} \cdots u_d^{k_d}.$$

(2) $K_n(\pi(u)) = u_1^{2k_1} \cdots u_d^{2k_d} - \frac{1}{\sqrt{n}} u_1^{k_1} \cdots u_d^{k_d} \xi_n(u)$. ここで，$\xi_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{a(X_i,u) - E_X[a(X,u)]\}$.

(3) $E[\xi_n(u)] = 0$,

(4) $E[\xi_n(u)\xi_n(v)]$
$= E_X[a(X,u)a(X,v)] - u_1^{k_1} \cdots u_d^{k_d} v_1^{k'_1} \cdots v_d^{k'_d}$

(5) ξ_n は，正規分布 ξ に法則収束し，$E[\xi(u)] = 0$,
$E[\xi(u)\xi(v)]$
$= E_X[a(X,u)a(X,v)] - u_1^{k_1} \cdots u_d^{k_d} v_1^{k'_1} \cdots v_d^{k'_d}$
が成り立つ.

4. 参考文献

[1] 舟木直久, 確率論, 朝倉書店, p.261, 2004.

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