P–4

学習理論と Vandermonde Matrix 型特異性

Learning theory and Vandermonde Matrix Type Singularities

○ 青柳 美輝 (Miki Aoyagi)[1]

概 要

In this paper, we consider Vandermonde Matrix Type Singularities. The log canonical threshold of such singularities corresponds to a learning coefficient of generalization error in Bayesian estimation, which serves to measure the learning efficiencies in hierarchical learning models [1].

1 導入

f を滑らかな実多様体 Y 上の恒等的に 0 でない実解析関数, $Z \subset Y$ を closed subscheme とする．log canonical threshold $\lambda_Z(Y,f)$ は解析的に $\lambda_Z(Y,f) = \sup\{c : |f|^{-c} \text{ is locally } L^1 \text{ near } Z\}$ で定義される [2]. f が多項式または収束級数の場合，複素数体上の log canonical threshold は，f の Bernstein-Sato 多項式 $b(s) \in \mathbb{C}[s]$ の最大根であることが知られている [3]. また，log canonical threshold $\lambda_Z(Y,f)$ はゼータ関数 $\int_{\text{near } Z} |f|^z \psi(w) dw$ の最大の極である．ここで，$\psi(w)$ は コンパクトサポートを持つ C^∞ 関数で，Z 上 $\psi(w) \neq 0$ を満たす．この論文では，Vandermonde matrix singularities の log canonical threshold の上限を与える．これは，学習理論における，混合正規分布，三層ニューラルネットワーク，混合二項分布のベイズ推測に関する学習効率を与える．異なるモデルから共通の特異性が現れたことから，学習理論において本質的ではないかと考えられている．

log canonical threshold は広中の特異点解消定理により，原理的には有限の手続きにより求められるが，具体的に求めるのは難しいとされている．計算機で代数計算により行う方式も提案されているが，Vandermonde matrix singularities は，パラメータを含んでいるため，確定された多項式の特異点解消よりも高度な面を含んでいる．更なる困難な問題点として，特異点が孤立していない・ニュートン図形が退化している等があげられる．

2 定義

* を付加した記号 a^*, b^*, w^* などは定数を表すものとする．

定義 1 U を w^* の十分小さな近傍とし，f を U 上の実解析関数，$\psi(w)$ をコンパクトサポートを持つ C^∞ 関数で，$\psi(w^*) \neq 0$ を満たすものとする．このとき，$\lambda_{w^*}(f)$ を $\int_U |f|^z \psi dw$ の最大の極とする．

定義 2 $Q \in \mathbb{N}$ を固定する． $b_1^* = \cdots = b_{i-1}^* = 0,\ b_i^* \neq 0$, かつ $\gamma_i = \begin{cases} 1 & \textit{if } Q \textit{ is odd,} \\ |b_i^*|/b_i^* & \textit{if } Q \textit{ is even.} \end{cases}$ のとき，$[b_1^*, b_2^*, \cdots, b_N^*]_Q = \gamma_i(0,\cdots,0,b_i^*,\cdots,b_N^*)$ と定義する．

また，

$$A_{M,H} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1H} \\ a_{21} & a_{22} & \cdots & a_{2H} \\ & & \vdots & \\ a_{M1} & a_{M2} & \cdots & a_{MH} \end{pmatrix}, B_{H,N,I} = \begin{pmatrix} \prod_{j=1}^N {b_{1j}}^{\ell_j} \\ \prod_{j=1}^N {b_{2j}}^{\ell_j} \\ \vdots \\ \prod_{j=1}^N {b_{Hj}}^{\ell_j} \end{pmatrix}, B_{H,N}^{(Q,m)} = (B_{H,N,I})_{\ell_1+\ldots+\ell_N=Qn+m, 0\leq n\leq H-1}.$$

$$(A_{M,H}, \mathbf{a}^*) = \begin{pmatrix} a_{11} & \cdots & a_{1H} & a_{1,H+1}^* \\ a_{21} & \cdots & a_{2H} & a_{2,H+1}^* \\ & \vdots & & \\ a_{M1} & \cdots & a_{MH} & a_{M,H+1}^* \end{pmatrix}, \begin{pmatrix} B_{H,N,I} \\ \mathbf{b}^* \end{pmatrix}, = \begin{pmatrix} \prod_{j=1}^N {b_{1j}}^{\ell_j} \\ \prod_{j=1}^N {b_{2j}}^{\ell_j} \\ \vdots \\ \prod_{j=1}^N {b_{Hj}}^{\ell_j} \\ \prod_{j=1}^N {b_{H+1,j}^*}^{\ell_j} \end{pmatrix}, \begin{pmatrix} B_{H,N}^{(Q,m)} \\ \mathbf{b}^* \end{pmatrix} = \begin{pmatrix} B_{H,N,I} \\ \mathbf{b}^* \end{pmatrix}_{\substack{\ell_1+\ldots+\ell_N=Qn+m, \\ 0\leq n\leq H}}$$

とする．ここで $\mathbf{a}^* = (a_{1,H+1}^*, \cdots, a_{M,H+1}^*)^t$, $\mathbf{b}^* = (b_{H+1,1}^*, \cdots, b_{H+1,N}^*)$ である．

[1]日大理工・教員・数学

定義 3 $Q \in \mathbb{N}$ を固定する．

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1H} & a^*_{1,H+1} & \cdots & a^*_{1,H+r} \\ a_{21} & \cdots & a_{2H} & a^*_{2,H+1} & \cdots & a^*_{2,H+r} \\ & \vdots & & & \vdots & \\ a_{M1} & \cdots & a_{MH} & a^*_{M,H+1} & \cdots & a^*_{M,H+r} \end{pmatrix}, \ I = (\ell_1, \ldots, \ell_N) \in {\mathbb{N}_{+0}}^N,$$

$$B_I = (\prod_{j=1}^N b_{1j}^{\ell_j}, \prod_{j=1}^N b_{2j}^{\ell_j}, \cdots, \prod_{j=1}^N b_{Hj}^{\ell_j}, \prod_{j=1}^N {b^*_{H+1,j}}^{\ell_j}, \cdots, \prod_{j=1}^N {b^*_{H+r,j}}^{\ell_j})^t,$$

および $B = (B_I)_{\ell_1+\cdots+\ell_N = Qn+1, n \geq 0} = (B_{(1,0,\cdots,0)}, B_{(0,1,\cdots,0)}, \cdots, B_{(0,0,\cdots,1)}, B_{(1+Q,0,\cdots,0)}, \cdots)$ とする．

a_{ki}，b_{ij} $(1 \leq k \leq M, 1 \leq i \leq H, 1 \leq j \leq N)$ は，定数 a^*_{ki}，b^*_{ij}，の近傍で定義された変数とする．

$\mathcal{I}$ を行列 AB の成分で生成されるイデアルとする．$\mathcal{I}$ で定められる特異点を **Vandermonde matrix 型特異点**とよぶ．簡単のため，$1 \leq j \leq r$ に対して，$(a^*_{1,H+j}, a^*_{2,H+j}, \cdots, a^*_{M,H+j})^t \neq 0, (b^*_{H+j,1}, b^*_{H+j,2}, \cdots, b^*_{H+j,N}) \neq 0$, および $j \neq j'$ に対して $[b^*_{H+j,1}, b^*_{H+j,2}, \cdots, b^*_{H+j,N}]_Q \neq [b^*_{H+j',1}, b^*_{H+j',2}, \cdots, b^*_{H+j',N}]_Q$ を仮定する．

3　Vandermonde matrix 型特異点の Log canonical threshold

定理 1 $(b^{**}_{01}, b^{**}_{02}, \cdots, b^{**}_{0N}) = (0, \ldots, 0)$ および，
$\{(b^{**}_{11}, \cdots, b^{**}_{1N}), \ldots, (b^{**}_{r'1}, \cdots, b^{**}_{r'N})$; $[b^*_{i1}, \cdots, b^*_{iN}]_Q \neq 0, i = 1, \ldots, H+r\}$ とする．特に $1 \leq i \leq r$ に対して，$(b^{**}_{i1}, \cdots, b^{**}_{iN}) = [b^*_{H+i,1}, \cdots, b^*_{H+i,N}]_Q$ とする．

$$\text{ここで，} [b^*_{i1}, \cdots, b^*_{iN}]_Q = \begin{cases} 0, & 1 \leq i \leq H_0 \\ (b^{**}_{11}, \cdots, b^{**}_{1N}), & H_0 + 1 \leq i \leq H_0 + H_1, \\ (b^{**}_{21}, \cdots, b^{**}_{2N}), & H_0 + H_1 + 1 \leq i \leq H_0 + H_1 + H_2, \\ \vdots & \\ (b^{**}_{r'1}, \cdots, b^{**}_{r'N}), & H_0 + \cdots + H_{r'-1} + 1 \leq i \leq H_0 + \cdots + H_{r'}, \end{cases}$$

かつ $H_0 + \cdots + H_{r'} = H$ であると仮定する．

このとき，$\alpha \geq 1$ に対して，${w_1^{(0)}}^* = \{a^*_{ki}, 0\}_{1\leq k\leq M, 1\leq i\leq H_0, 1\leq j\leq N}$, ${w_1^{(\alpha)}}^* = \{a^*_{k,H_0+\cdots+H_{\alpha-1}+i}, 0\}_{1\leq k\leq M, 2\leq i\leq H_\alpha, 1\leq j\leq N}$, $\mathbf{a}^{(\alpha)^*} = (a^*_{1,H+\alpha}, \cdots, a^*_{M,H+\alpha})^t$ とおくと

$$\lambda_{w^*}(||AB||^2) = \frac{Mr'}{2} + \lambda_{w_1^{(0)*}}(||A_{M,H_0} B^{(Q,m)}_{H_0,N}||^2)$$
$$+ \sum_{\alpha=1}^{r} \lambda_{w_1^{(\alpha)*}}(||(A_{M,H_\alpha-1}, \mathbf{a}^{(\alpha)^*}) B^{(1,1)}_{H_\alpha,N}||^2) + \sum_{\alpha=r+1}^{r'} \lambda_{w_1^{(\alpha)*}}(||A_{M,H_\alpha-1} B^{(1,1)}_{H_\alpha-1,N}||^2).$$

が成立する．

定理 2　次の上限が得られた．

$$\lambda_0(||A_{M,H} B^{(Q,m)}_{H,N}||^2), \lambda_0(||(A_{M,H-1}, \mathbf{a}^*) B^{(Q,m)}_{H,N}||^2) \leq \begin{cases} \dfrac{NH}{2m}, & \textit{if } N \leq mM \leq m(N-1), \\ \dfrac{NH}{2m}, & \textit{if } M \geq N, (N-1)(m-1) \geq 1, \\ \dfrac{2HN + Q(M(1+k) + (N-1)(2H-k-1))k}{4Qk + 4m}, & \\ \quad \textit{if } M \geq N, (N-1)(m-1) = 0, & \end{cases}$$

$$\lambda_0(||A_{M,H} B^{(Q,m)}_{H,N}||^2) \leq \frac{MH}{2}, \quad \lambda_0(||(A_{M,H-1}, \mathbf{a}^*) B^{(Q,m)}_{H,N}||^2) \leq \frac{mM(H-1)+N}{2m}, \textit{ if } mM \leq N-1.$$

ここで，$k = \max\{i \in \mathbb{Z}; 2H \geq (Qi(i-1) + 2mi)(M-N+1)\}$.

この上限は真の値と予想している．

参考文献

[1] S. Watanabe. Algebraic geometrical methods for hierarchical learning machines. *Neural Networks*, 14(8):1049–1060, 2001b.

[2] J. Kollár. Singularities of pairs. *Algebraic geometry-Santa Cruz 1995, Proc. Sympos. Pure Math., Amedsr. Math. Soc., Providence, RI,*, 62:221–287, 1997.

[3] M. Kashiwara. B-functions and holonomic systems. *Inventions Math.*, 38:33–53, 1976.