P-13

学習理論における特異性の超関数を使った解析について
Analysis of singularities in learning theory by Schwartz distribution

○高橋健彰 [1], 青柳美輝 [2]
Takeaki Takahashi, Miki Aoyagi

Abstract: The purpose of a learning system is to estimate an unknown true density function that generates the data. In order to analyze such data, we usually use hierarchical learning models. These are, however, singular learning models, which cannot be analyzed using the classic theory of regular statistical models. In this paper, we discuss the method to analyze singularities in learning theory, using a Schwartz distribution. The Mellin transform of the Schwartz distribution is a zeta function in learning theory. The poles of such zeta functions give the behavior of stochastic complexities, which are likelihood functions of learning models in Bayesian estimation.

1. 導入

機械の文字認識，画像認識，音声認識，遺伝子解析などにおいては，計測して得られた多量のデータから，そのデータを発している情報源の確率分布を推測する必要がある．情報分野では，この推測を学習といい，その学習の仕組みをまとめて体系化したものを「学習理論」と呼んでいる．例えば，文字認識の場合，手書き文字は，ノイズの含まれたデータとみて，確率的に考える必要性がある．このような文字認識における情報源の確率分布は，正規分布に従うような単純なものではない．多くのデータから，機械が認識出来るように訓練する機械学習においては，複雑な確率分布を表現できる階層構造・内部構造をもつ神経回路網，混合正規分布や縮小ランクなどが利用されている．学習理論は，これらを利用した場合の学習の挙動を定めている共通または固有の法則を解明し，その上で情報科学的なシステム設計法を与えることを目的としている．近年，この研究のために，解析関数論の手法をもちいた理論研究が盛んになってきた．この論文では，その解析結果の一部を紹介する．

2. 超関数

定義 1 φ を $\mathbb{R}^d$ から $\mathbb{C}$ へ関数とする．φ が，次を満たすとき C_0^∞ 級関数という．

(1) $\varphi(w)$ が C^∞ 級関数.

(2) $\mathrm{supp}\varphi=\overline{\{w\in\mathbb{R}^d;\varphi(w)\neq 0\}}$ がコンパクト.

また，C_0^∞ の位相を次で与える．次の (3), (4) が成立する時，$C_0^\infty \ni \varphi_k \to \varphi \in C_0^\infty (k\to\infty)$ であるとする．

(3) コンパクト $K\subset \mathbf{R}^d$ が存在して，

$$\mathrm{supp}\varphi\subset K,\quad \mathrm{supp}\varphi_{\mathrm{k}}\subset K$$

(4) $\alpha=(\alpha_1,\cdots,\alpha_d)\in(\mathbb{N}\cup\{0\})^d$ に対して

$$\lim_{k\to\infty}\max_{w\in K}|\partial_\alpha\varphi_k(w)-\partial_\alpha\varphi(w)|=0.$$

定義 2 C_0^∞ から $\mathbb{C}$ への関数 T が，線形および連続であるとき $\mathbb{R}^d$ 上の超関数であるという．

定義 3 f を開集合 $U\subset\mathbb{R}^d$ から $\mathbb{R}$ への解析関数とする．超関数 $\delta(t-f(x))$ を次で定義する．

$f(x)=t$ のとき，$\nabla f(x)\neq 0$ であると仮定する．

$\{U_j\}_{j=1}^\infty$ を U の十分小さな開集合の被覆とする．

$U_j\cap\{f(x)=t\}$ 上では，ある $1\leqq i_j\leqq d$ に対して，$\frac{\partial}{\partial x_{i_j}}f\neq 0$ となるようにとっておく．$\{U_j\}_{j=1}^\infty$ の 1 の分解を $\{\rho_j(x)\}$ とする．このとき，$(x_1,\cdots,x_{i_j},\cdots,x_d)$ から $(u_1,\cdots,f,\cdots,u_d)$ へ変数変換を行い，U にサポートを持つ $\varphi\in C_0^\infty$ に対して，

$$\int\delta(t-f(x))\varphi dx = \sum_{U_j}\int\frac{\varphi\,\rho_j}{|\frac{\partial}{\partial x_{i_j}}f|}(u_1,\cdots,t,\cdots,u_d)\, du_1\cdots du_{i_j-1}du_{j_i+1}du_d$$

と定義する．

定理 4 $v(t)=\int\delta(t-f(x))\varphi(x)dx$ とおく．$\mathbb{R}^1$ から $\mathbb{R}^1$ への局所積分可能な関数 F に対し，

$$\int_{\mathbb{R}^1}F(t)v(t)dt=\int_{\mathbb{R}^d}F(f(x))\varphi(x)dx.$$

定義 5 可測関数 $f:(0,\infty)\to\mathbb{C}$ に対して，$F(z)=\int_0^\infty f(t)t^z dt$ が $|F(z)|<\infty$ を満たすとき，$F(z)$ を f のメリン変換という．

定理 6 (メリン逆変換) $a<\mathrm{Re}(z)<b$ において，

$$\int_0^\infty|f(t)|t^{\mathrm{Re}(z)}dt<\infty$$

かつ $f(t)$ が $t=s$ の近傍で有界変動ならば，$a<c<b$ において，

$$\frac{1}{2}\{f(s+0)+f(s-0)\}=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}F(z)s^{-z-1}dz$$

1: 日大理工・院 (前)・数学　2: 日大理工・教員・数学

がなりたつ

定理 7 a, b を正の実数，$k_1, \cdots, k_r$ を自然数, $h_1, \cdots, h_r$ を 0 以上の整数とする．ここで $\lambda = \frac{h_1+1}{2k_1} = \cdots = \frac{h_r+1}{2k_r}$ が成立すると仮定する．このとき

$$v(t) = \int_{[0,b]^r} \delta(t - ax_1^{2k_1} \cdots x_r^{2k_r}) x_1^{h_1} \cdots x_r^{h_r} dx_1 \cdots dx_r$$

であるならば,

$$v(t) = \begin{cases} \gamma_b \frac{t^{\lambda-1}}{a^\lambda} \left(\log \frac{ab^{2|k|}}{t}\right)^{r-1} & (0 < t < ab^{2|k|}) \\ 0, & \text{その他} \end{cases}$$

が成立する．ここで $|k| = k_1 + \cdots + k_r$, $|h| = h_1 + \cdots + h_r$ および $\gamma_b = \frac{b^{|h|+r-2(|k|)\lambda}}{2^r(r-1)!k_1k_2\cdots k_r}$.

3. 学習理論

$x \in \mathbb{R}^N$，$w \in W \subset \mathbb{R}^d$ 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

$$p(w|D_n) = \frac{1}{Z_n(X_1, \cdots, X_n)} \psi(w) \prod_{i=1}^{L} p(X_i|w)^\beta$$

で与えられる．ここで，

$Z_n(X_1, \cdots, X_n) = \int \varphi(w) \prod_{i=1}^n p(X_i|w)^\beta dw$

である．特に，$\beta = 1$ のとき，

$\int Z_n(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1$

となる．すなわち，与えられた学習モデル $p(x|w)$ と事前分布 $\varphi(w)$ に対して，D_n の確率分布を表している．したがって，$p(x|w)$ と $\varphi(w)$ の尤度関数ともいえる．$F_n = -\log Z_n$ を確率的複雑さという．カルバック距離を $K(w) = \int q(x) \log \frac{q(x)}{p(x|w)} dx$, とおく. 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. このとき,

$F_n^0 = -\log(\int \varphi(w) \exp(-n\beta K_n(w)) dw)$

に対して，$F_n = F_n^0 + \beta \sum_{i=1}^n \log q(X_i)$ が成り立つ.

定理 8 $\zeta(z) = \int K(w)^z \varphi(w) dw$ とおく．この有理型関数の極は，すべて，実数であり負の有理数である．それを順に $0 > -\lambda_1 > -\lambda_2 > -\lambda_3 > \cdots$ とおく．m_k を λ_k の位数とする.

このとき,

$$v(t) = \int \delta(t - K(w)) \varphi(w) dw$$

$$= \sum_{k=1}^{\infty} \sum_{m=1}^{m_k} c'_{km} t^{\lambda_k - 1} (-\log t)^{m-1}$$

が成り立つ.

(証明概略)

特異点解消定理 (Hironaka Theorem[1]) 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, Y_w から U_w への写像 π が存在して，Y_w の局所座標 $(u_1, u_2, \cdots, u_d)$ において，$K(\pi(u)) = u_1^{2k_1} \cdots u_d^{2k_d}$ と書ける．よって，ゼータ関数 $\zeta(z) = \int K(w)^z \psi(w) dw$ の U_w での局所的な積分は

$$\begin{aligned} \zeta_w(z) &= \int_{U_w} K(w)^z \psi(w) \mathrm{d}w \\ &= \int_{Y_w} (u_1^{2k_1} u_2^{2k_2} \cdots u_d^{2k_d})^z \psi(\pi(y)) Jac_\pi(u) \mathrm{d}u \end{aligned}$$

となる．この積分は初等的に求めることができる．すなわち $\zeta(z)$ の極，位数を得ることができる.

よって，ある正則関数 $\zeta_0(z)$ が存在して，

$$\zeta(z) = \zeta_0(z) + \sum_{k=1}^{\infty} \sum_{m=1}^{m_k} \frac{c_{km}}{(z + \lambda_k)^{m_k}}$$

となる.

また，$v(t) = \int \delta(t - K(w)) \varphi(w) dw$ は，$\zeta(t) = \int_0^\infty v(t) t^z dt$ を満たす．定理 4, 6, 7 を適用することによって，$t \to 0$ における漸近展開

$$v(t) = \sum_{k=1}^{\infty} \sum_{m=1}^{m_k} c'_{km} t^{\lambda_k - 1} (-\log t)^{m-1}$$

を得る.

定理 9 (渡邊 [3])

$$F_n^0 = \lambda \log n - (m-1) \log \log n + F^R(\xi) + o_p(1)$$

ここで，$\lambda = \lambda_1$ であり，m は λ の最大の位数である．$F^R(\xi)$ は確率変数，$o_p(1)$ は，0 に収束する確率変数である.

(証明概略) $v(t)$ のラプラス変換 [2] は，$Z(n) = \int \exp(-nK(w)) \varphi(w) dw$ であること，および $Z_n = \int \exp(-nK_n(w)) \varphi(w) dw$ を用いて証明する.

4. 参考文献

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

[2] 竹之内 脩: フーリエ展開, 秀潤社, p.239, 1981.

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