2018年10月25日

Watanebe, 6.3

@unaoya

統計的学習

1. 真の分布$q(x)$またはそこから生成されるサンプルを予測したい
2. モデル$p(x|w)$とパラメータ空間Wを設定
3. 与えられたサンプルからパラメータ空間上の測度および予測分布$\hat{p}(x)$を決定

モデルを評価したい

Bayes quartet

四つのモデルの評価基準

	Bayes推測	Gibbs推測
分布	B_g	G_g
サンプル	B_t	G_t

これらはサンプルD_nに依存した確率変数

6.3の目標

- サンプル数$n \to \infty$における漸近挙動
- サンプルについての期待値
- これらの4つの間の関係

について調べる。

今日はG_gについて

Gibbs推測

事後分布に従ってパラメータ $\hat{w}$ をサンプリングし、$\hat{p}(x) = p(x|\hat{w})$ を予測分布とする。

汎化誤差 G_g

$q(x)$ と $\hat{p}(x)$ の KL divergence を w について事後分布 $p(w|D_n)$ で積分したもの

$$G_g = \int_W K(w)p(w|D_n)dw$$

問題

$n \to \infty$でnG_gがどのような確率変数に収束するか？

主要項

$$G_g(\epsilon) = \frac{\int_{K(w)\le\epsilon} K(w)p(w|D_n)dw}{\int_{K(w)\le\epsilon} p(w|D_n)dw}$$

補題 1 (Lemma 6.3). $nG_g - nG_g(\epsilon)$は0に確率収束する。

主要項の評価

$$G_g(\epsilon) = \frac{\int_{K(w)\le\epsilon} K(w)p(w|D_n)dw}{\int_{K(w)\le\epsilon} p(w|D_n)dw}$$
$$= E[K(w)|_{K(w)\le\epsilon}]$$

の評価をしたいが直接は難しい。

特異点解消を使う

標準形

$$f(x, g(u)) = \log(\tfrac{q(x)}{p(x|g(u))}) = a(x, u)u^k$$ とし

$$K(g(u)) = u^{2k}$$

$$K_n(g(u)) = u^{2k} - \frac{1}{\sqrt{n}}u^k \xi_n(u)$$

$$\xi_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{a(X_i, u) - E_X[a(X, u)]\}$$

$\xi_n(u)$ はサンプル D_n に依存した確率過程。ξ_n は Gauss 過程 ξ に法則収束する。

補題 2 (6.51).

$$G_g^*(\xi_n) = E_{y,t}[t|\xi_n]$$

と定義すると

$$nG_g(\epsilon) - G_g^*(\xi_n) \to^P 0$$

特異点解消 M 上での積分 $E_{y,t}$ と E_u

- $\xi(u)$: M 上の C^1 級関数（サンプルの確率過程）
- $f(u)$: M 上の関数（$K(w)$ に対し $f(u)=u^{2k}$）
- $0 \leq \sigma \leq 1$

$$E_u^\sigma[f(u)|\xi] = \frac{\displaystyle\sum_{\alpha\in A}\int_{[0,b]^d} f(u)Z(u,\xi)du}{\displaystyle\sum_{\alpha\in A}\int_{[0,b]^d} Z(u,\xi)du}$$

A は座標近傍の（有限）集合。

$Z(u,\xi)$ は

$$u^h\phi^*(u)\exp(-\beta n u^{2k}+\beta\sqrt{n}u^k\xi(u)-\sigma u^k a(X,u))$$

事後分布 $p(w|D_n)$ との関係。

- $u^h\phi^*(u)$ が事前分布 $\phi(w)$ に対応。
- $\sigma=0$ として

$$\begin{aligned} Z_n^0 p(w|D_n) &= \exp(-n\beta K_n(w)) \\ &= \exp(-\beta n u^{2k} + \beta\sqrt{n}u^k\xi_n(u)) \end{aligned}$$

補題 3 (6.41).

$$G_g(\epsilon) = E_u^0[u^{2k}|\xi_n]$$

$u^{2k} = K(g(u))$であった。

本質的部分

座標 $u=(x,y)$ **と本質的部分** $A^* \subset A$ **（**$K(w)$ **の特異点解消から決まる）**

$E_{y,t}[f(y,t)|\xi]=$

$$\frac{\displaystyle\sum_{\alpha\in A^*}\int dt\int_{[0,b]^{d-m}} f(y,t)Z_0(y,t,\xi)du}{\displaystyle\sum_{\alpha\in A^*}\int dt\int_{[0,b]^{d-m}} Z_0(y,t,\xi)du}$$

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

補題 4 (Lemma 6.6, $p = 1, f = 1, \xi = \xi_n$)**.**

$$|E_u^0[nu^{2k}|\xi_n] - E_{y,t}[t|\xi_n]| \le \frac{D(\xi_n, 1, \phi^*)}{\log n}$$

これの証明に４章での分配関数の計算を用いる。

収束先の構成

定義 1 (6.46). M 上の関数 $\psi(u)$ に対し

$$G_g^*(\psi) = E_{y,t}[t|\psi]$$

これを使ってさっきの補題を書き直すと

補題 5.

$$|nG_g(\epsilon) - G_g^*(\xi_n)| \leq \frac{D(\xi_n, 1, \phi^*)}{\log n}$$

結論

ξ_n が ξ に法則収束することから

- 補題4を用いて $nG_g(\epsilon) - G_g^*(\xi_n) \to 0$
- $G_g^*(\xi_n) - G_g^*(\xi) \to 0$

が言える。

全て合わせて $nG_g - G_g^*(\xi) \to 0$ が証明できた。

$G_g^*(\xi)$ は ξ に依存した確率変数である。

4章の復習

ゼータ関数

開集合 $U \subset \mathbb{R}^d$ 上の非負解析的関数 $K(w)$ とコンパクト台 C^∞ 関数 $\phi(w)$ に対し、

$$\zeta(z) = \int K(w)^z \phi(w) dw$$

と定義する。これの極の位置とその位数はどのような情報を持つか？

状態密度関数

$\zeta(z)$ の逆 Mellin 変換は状態密度関数

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

である。Mellin 変換の理論により、これの発散のオーダーが $\zeta(z)$ の極の位置と対応。

分配関数

状態密度関数$v(t)$のLaplace変換

$$Z(n) = \int \exp(-nK(w))\phi(w)dw$$

を分配関数という。これが6章前半で調べていたもの。

特異学習理論

Remark 4.4にあるように

$$Z = \int \exp(-n\beta K(w) + \beta\sqrt{nK(w)}\xi(w))\phi(w)dw$$

の$n \to \infty$での挙動を調べたい。

K についての特異点解消により、normal crossing の場合の積分 $Z(n, \xi, \phi)$ を用いて

$$Z = \sum_{\alpha} Z(n, \xi \circ g_\alpha, \phi \circ g_\alpha |g'_\alpha|)$$

と書けるので、$Z(n, \xi, \phi)$ について調べるのが 4.4 の目標。

$$Z^p(n,\xi,\phi) =$$

$$\int_{[0,b]^r} dx \int_{[0,b]^s} dy K(X,y)^p x^h y^{h'} \phi(x,y)$$

$$\exp(-n\beta K(x,y)^2 + \sqrt{n}\beta K(x,y)\xi(x,y))$$

と定義する。

さらにこれで$\xi = 0, \phi = 1$と置いたものを

$$Z^p(n) = \int_{[0,b]^r} dx \int_{[0,r]^s} dy$$
$$K(x,y)^p x^h, y^{h'} \exp(-n\beta K(x,y)^2)$$

と書くことにする。

定理 1 (Theorem 4.7).

$$\frac{h_i + 1}{2k_i} = \lambda$$

が一定で

$$\frac{h'_j + 1}{2k'_j} > \lambda$$

とする。$K(x, y) = x^k y^{k'}$ のときに、ある $a_1, a_2 > 0$ が存在して任意の n に対して

$$a_1 \frac{(\log n)^{r-1}}{n^{\lambda+p}} \leq Z^p(n) \leq a_2 \frac{(\log n)^{r-1}}{n^{\lambda+p}}$$

平均誤差関数 $K(w)$ と事前分布 $\phi(w)$ に対してゼータ関数

$$\zeta(z) = \int K(w)^z \phi(w) dw$$

を定義する。これの極の情報から K の？特異点の実対数閾値がもとまる。これが自由エネルギーおよび汎化損失の理論値が明らかになる。

0.1 疑問点

正則な場合の計算をやる。特にこの時正則性をどこで使うか。正則な場合、サンプルが多いことが仮定される？