Mathematical Foundations of Infinitie-Dimensional Statistical Model

Chap.3.5.3 - 3.6.1

Evarist Gine, Richard Nickl

Presenter: yc, Choi

2019.01.04

Contents

Contents

Reviews

- Theorem 3.1.7(Bernstein's inequality) Let $X_i, (1 \leq i \leq n)$ be centred independent random variables such that, for all $k \geq 2$ and all $1 \leq i \leq n$),

$$E|X_i|^k \leq \frac{k!}{2}\sigma_i^2 c^{k-2}$$

and set $\sigma = \sum_{i=1}^{n} \sigma_i$, $S_n = \sum_{i=1}^{n} X_i$. Then

$$Pr\{S_n \geq t\} \leq \exp(-\frac{t^2}{2\sigma + 2ct}), t \geq 0$$

Reviews

- Theorem 3.5.13 Let P be a probability measure on $(S, \mathcal{S})$ and for any $n \in \mathbb{N}$, and let $X_1, \ldots, X_n$ be an independent sample of size n from P. Let $\mathcal{F}$ be a class of measurable functions on S that admits a P-square integrable envelope F and satisfies the $L^2(P)$-bracketing condition

$$\int_0^2 \sqrt{\log(N_{[]}(\mathcal{F}, L^2(P), \|F\|_{L^2(P)}\tau)}d\tau < \infty$$

Set $\sigma^2 := \sup_{f \in \mathcal{F}} Pf^2$ and

$$a(\delta) = \frac{\delta}{\sqrt{32 log(2(N_{[]}(\mathcal{F}, L^2(P), \delta/2)}}$$

Then, for any $\delta > 0$,

$$E\|\sum_{i=1}^{n}(f(X_i) - Pf)\|^*_{\mathcal{F}} \leq 56\sqrt{n}\int_0^{2\delta}\sqrt{2\log(N_{[]}(\mathcal{F}, L^2(P), \tau)}d\tau + 4nP[FI(F > \sqrt{n}a(\delta)] + \sqrt{n\sigma$$

- Instead of the $L^2(P)$-norm, we will use a quantity $\rho_K(f)$ that is neither a norm nor a distance.
- Given a probability measure P on $(S, \mathcal{S})$ and a positive constant K, we set

$$\rho_K^2(f) = 2K^2 E(e^{|f(X)|/K} - 1 - |f(X)|/K) = 2K^2 \sum_{k=2}^{\infty} \frac{E|f(X)|^k}{K^k k!}$$

  where X has probability law P
- Define the Bernstein size of f as the nonnegative square root of $\rho_K^2(f)$.

Properties of $\rho_K(f)$

- $\rho_K(f) < \infty$ if and only if $Ee^{|f(x)|/K} < \infty$
- $\lim_{K\to\infty} \rho_K(f) = Ef^2(X)$
- Lemma 3.5.18
  (a) $\rho_K(f)$ is nonincreasing in K and $\rho_K(\lambda f) = |\lambda|\rho_{K/|\lambda|}(f)$
  (b) $\lambda_K^2(f+g) \leq 2\rho_{K/2}^2(f) + 2\rho_{K/2}^2(g)$
  (c) If $\rho_K(f) \leq R$, then $E|f(X)|^k \leq k!K^{k-2}R^2/2$, for all $k \geq 2$.
  (d) If $E|f(X)|^k \leq k!K^{k-2}R^2/2$, for all $k \geq 2$, then $\rho_{2K}^2(f) \leq 2R^2$
  (e) If $\|f\|_\infty \leq K$ and $Ef(X)^2 \leq R^2$, then $\rho_{2K}^2 \leq 2R^2$

- Corollary 3.5.19 Let $X_i$be independent identically distributted random variables with law P, and let P_n be the corresponding empirical measure. Assume that $Ef(X) = 0$ and $\rho_K(f) \leq R$. Then, given C>0,

$$Pr\{\sqrt{n}|(P_n - P)(f)| \geq t\} \leq 2\exp(-\frac{t^2}{2(C+1)R^2}), \quad t \leq C\sqrt{n}R^2/K$$

- Definition 3.5.20 Let $\mathcal{F}$ be a class of measurable functions $f\colon S \to \mathbb{R}$ such that $Ee^{|f(X)|/K} < \infty$. For each $\epsilon > 0$, the $B(K,P)$-bracketing number $N_{BK}(\mathcal{F}, P, \epsilon)$ is defined as the smallest N for which there exists a partition of the class $\mathcal{F}$ into N subsets $B_1, \dots, B_N$ such that, letting $\Delta_i := (\sup_{f,g \in B_i} |f - g|)^*$, we have $\rho_K(\Delta_i) \le \epsilon$, for $1 \le i \le N$. For each $\epsilon > 0$, the $B(K,P)$-bracketing entropy of $\mathcal{F}$ is defined as $H_{BK}(\mathcal{F}, P, \epsilon) = \log N_{BK}(\mathcal{F}, P, \epsilon)$.

- Theorem 3.5.21 Let P be a probability measure on $(S, \mathcal{S})$, and for each n, let P_n be the empirical measure corresponding to n independent identically distributed random variables with law P. Let $\mathcal{F}$ be a class of measurable functions such that $\rho_K(f) \leq R$, for all $f \in \mathcal{F}$. Given $C_1 < \infty$, for all C sufficiently large and C_0 satisfying

$$C_0^2 \geq C^2(C_1 + 1)$$

, and for $n \in \mathbb{N}$ and $t > 0$ satisfying

$$C_0(R \vee \int_{t/(2^6\sqrt{n}}^{R} \sqrt{H_{B,K}(\mathcal{F}, P, \epsilon) d\epsilon}) \leq t \leq \sqrt{n}((8R) \wedge (C_1 R^2/K))$$

we have

$$Pr\{\sqrt{n}\|P_n - P\|_{\mathcal{F}} \geq t\} \leq C \exp(-\frac{t^2}{C^2(C_1 + 1)R^2}$$

Contents

VC-class

Let $\mathcal{C}$ be a class of subsets of a set S. Let $A \subseteq S$ be a finite set.

- The trace of $\mathcal{C}$ on A is the collection of all the subsets of A obtained by intersection of A with sets C from $\mathcal{C}$.
- $\Delta^{\mathcal{C}}(A)$ is the cardinality of the trace of the class $\mathcal{C}$ on A.
- $\Delta^{\mathcal{C}}(A) \leq 2^{Card(A)}$
- $\mathcal{C}$ *shatters* A if $\Delta^{\mathcal{C}}(A) = 2^{Card(A)}$
- $m^{\mathcal{C}}(k) = \sup_{A \subseteq S; Card(A)=k} \Delta^{\mathcal{C}}(A)$.

VC-class

Let $\mathcal{C}$ be a class of subsets of a set S. Let $A \subseteq S$ be a finite set.

- Definition 3.6.1 A collection of sets $\mathcal{C}$ is a Vapnik-Červonenkis Classes if quantity

$$v(\mathcal{C}) := \begin{cases} min\{k : m^{\mathcal{C}}(k) < 2^k\} & \text{if } m^{\mathcal{C}}(k) < 2^k \\ \infty & \text{otherwise.} \end{cases}$$

is finite, that is, if there exists $k < \infty$ such that $\mathcal{C}$ does not shatter any subsets of S of cardinality k.

VC-class

- Theorem 3.6.2 Let $\mathcal{C}$ be a non-empty VC class, and let $v = v(\mathcal{C})$. Then, for any finite set $A \subseteq S$,

$$\Delta^{\mathcal{C}}(A) \leq Card\{B \subseteq A : Card(B) < v\} = \sum_{j=0}^{v-1} \binom{Card(A)}{j}$$

and therefore,

$$m^{\mathcal{C}}(n) \leq \sum_{j=0}^{v-1} \binom{n}{j}$$

VC-class

- Proposition 3.6.4 Let k and n be nonnegative integers such that $n \geq k+2$. Then

$$\sum_{j=0}^{k} \binom{n}{j} \leq \frac{1.5n^k}{k!}$$

- Corollary 3.6.5 If $\mathcal{C}$ is a non-empty VC class of sets and $v = v(\mathcal{C})$ is its VC index, then,

$$m^{\mathcal{C}}(n) \leq \frac{1.5n^{v-1}}{(v-1)!}, \text{ for} n \geq v+1$$

For $n = v$, $m^{\mathcal{C}}(n) \leq 2^v - 1$, and for $n < v$, $m^{\mathcal{C}} \leq 2^n < 2^v - 1$. In particular,

$$m^{\mathcal{C}}(n) \leq 2n^{v-1}, \text{ for all} n \geq 1.$$

VC-class

- Proposition 3.6.6 If $\mathcal{G}$ is a finite-dimensional vector space of real functions on S, then the class of sets $C := [\{g \geq 0\} : g \in \mathcal{G}]$is VC with $v(C) = \mathrm{dim}\mathcal{G} + 1$. The same is true for $[\{g > 0\} : g \in \mathcal{G}]$
- Proposition 3.6.7

  (i) If $\mathcal{C}$ is VC, then $\mathcal{C}^c := \{C^c : C \in \mathcal{C}\}$ is VC.

  (ii) If $\mathcal{C}$ and $\mathcal{D}$ are VC, Then $\mathcal{C} \cup \mathcal{D}$and $\mathcal{C} \cap \mathcal{D}$ are VC.

  (iii) If $\mathcal{C}$ is a collection of subsets of S and $\mathcal{D}$ is a collections of subsets of T and both are VC, then $\mathcal{C} \times \mathcal{D}$ is also VC.

  (iv) If $\mathcal{C} \subset \mathcal{D}$ and $\mathcal{D}$ are VC, then $\mathcal{C}$ is VC.