Mathematical Foundations of Infinite-Dimensional Statistical Models

Chap. 3.6.2 - 3.6.3

Evarist Giné, Richard Nickl

Presenter: Sarah Kim

2019.01.11

Contents

Reviews

- Notations
  - $\mathcal{C}$: a class of subsets of a set S, $A \subseteq S$: a finite set;
  - trace($\mathcal{C}$) on $A = \{A \cap C : C \in \mathcal{C}\}$;
  - $\Delta^{\mathcal{C}}(A)$: the cardinality of trace of the class $\mathcal{C}$ on A;
- Let $m^{\mathcal{C}}(k) = \sup\limits_{\substack{A \subseteq S \\ \text{Card(A)}=k}} \Delta^{\mathcal{C}}(A)$.
- **Definition 3.6.1** $\mathcal{C}$ is VC class if

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

is finite.

VC Subgraph

- **Definition 3.6.8** The subgraph of a real function f on S is the set

$$G_f = \{(s,t) : s \in S, t \in \mathbb{R}, t \leq f(s)\}.$$

A class of functions $\mathcal{F}$ is *VC subgraph* of index ν if the class of sets $\mathcal{C} = \{G_f : f \in \mathcal{F}\}$ is VC of index ν.

VC Subgraph

- **Examples**
  1. Suppose $\mathcal{C}$ is a VC class of index $\nu(\mathcal{C})$, then $\mathcal{F} := \{\mathbf{I}_C : C \in \mathcal{C}\}$ is VC subgraph of index $\nu(\mathcal{C})$.
  2. (Lemma 2.6.15, VW1996) Any finite-dimensional vector space $\mathcal{F}$ of measurable functions $f : S \to \mathbb{R}$ is VC subgraph of index $\leq \dim(\mathcal{F}) + 1$.

- Main result (Thm 3.6.9) shows that the $L^p(P)$-covering numbers of $\mathcal{F}$ admit small bounds, of the order of $\epsilon^{-(\nu-1)p}$, uniformly in P.

- **Theorem 3.6.9 (Dudley-Pollard)** Let $\mathcal{F}$ be a non-empty VC subgraph class of functions admitting an envelope $F \in L^p(S, \mathcal{S}, P)$ for some $1 \leq p < \infty$. Suppose that the class $\mathcal{C}$ of subgraphs of the function in $\mathcal{F}$ has index ν. Set $m_{\nu,w} = \max\{m \in \mathbb{N} : \log m \geq m^{1/(\nu-1)-1/w}\}$ for $w > \nu - 1$. Then

$$D(\mathcal{F}, L^p(P), \epsilon\|F\|_p) \leq m_{\nu,w} \vee \left[2^{w/(\nu-1)} \left(\frac{2^{p+1}}{\epsilon^p}\right)^w\right], \text{ for all } w > v - 1. \tag{3.233}$$

Proof of Theorem 3.6.9

- Let $f_1, \ldots, f_m$ be a maximal collection of functions in $\mathcal{F}$ with

$$P|f_i - f_j|^p > \epsilon^p PF^p, \quad i \neq j,$$

  so that $m = D(\mathcal{F}, L^p(P), \epsilon\|F\|_p)$ and let C_i be the subgraph of f_i.

- We calculate the probability that at least two graphs have same intersection with sample $\{(s_j, t_j), j = 1, \ldots, k\}$ and let k be s.t. this probability is less than 1. Then there exists a set of $k(\leq (2^{p+1}/\epsilon^p)\log m)$ sample s.t. $C_i \in \mathcal{C}, i = 1, \ldots, m$ intersect different subsets of this set, hence $m^{\mathcal{C}}(k) \geq m$.

- By Cor. 3.6.5,

$$m \leq m^{\mathcal{C}}(k) \leq 2k^{\nu-1} \leq 2\left(\frac{2^{p+1}}{\epsilon^p}\log m\right)^{\nu-1},$$

  and some algebra gives the desired bound.

Application of Theorem 3.6.9

- $$\sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i) - Pf)$$
- Using Theorem 3.5.1, 3.5.4, if $0 \in \mathcal{F}$,

$$\begin{aligned} E\sqrt{n}\|P_n - P\|_{\mathcal{F}} &\leq 8\sqrt{2}E\left[\int_0^{\sqrt{\|P_n f^2\|_{\mathcal{F}}}} \sqrt{\log 2D(\mathcal{F}, L^2(P_n), \tau)}d\tau\right] \\ &\leq 8\sqrt{2}\|F\|_{L^2(P)}\int_0^1 \sup_{Q:\text{finitely discrete}} \sqrt{\log 2D(\mathcal{F}, L^2(Q), \epsilon\|F\|_{L^2(Q)})}d\epsilon \\ &\lesssim \|F\|_{L^2(P)}\int_0^1 \sqrt{\nu\log(A/\epsilon)}d\epsilon, \end{aligned}$$

where A only depends on ν.

VC type

- By Thm. 3.6.9, if $\mathcal{F}$ is VC subgraph, then, for any probability measure P on $(S, \mathcal{S})$,

$$N(\mathcal{F}, L^2(P), \epsilon \|F\|_{L^2(P)}) \leq \left(\frac{A}{\epsilon}\right)^{2w},$$

  where $w > \nu$ and A depending on ν, w.

- **Definition 3.6.10** A class of measurable functions is of VC type w.r.t. measurable envelope F of $\mathcal{F}$ if there exist fininte constants A, w s.t. for all probability measures Q on $(S, \mathcal{S})$

$$N(\mathcal{F}, L^2(Q), \epsilon \|F\|_{L^2(Q)}) \leq (A/\epsilon)^w.$$

- Given $1 \leq p < \infty$, a function $f : \mathbb{R} \mapsto \mathbb{R}$ is of *bounded p-variation* if the quantity

$$v_p(f) := \sup \left\{ \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|^p : -\infty < x_0 < \cdots < x_n < \infty, n \in \mathbb{N} \right\}$$

  is finite.

- **Proposition 3.6.12** Let f be a function of bounded p-variation, $p \geq 1$. Then the collection $\mathcal{F}$ of translations and dilations of f

$$\mathcal{F} = \{x \mapsto f(tx - s) : t > 0, s \in \mathbb{R}\}$$

  is of VC type.

Expansion to Density Estimation

- In the density estimation based on convolution kernels, the corresponding class of functions is

$$\mathcal{K} = \{K((t - \cdot)/h) : t \in \mathbb{R}, h > 0\},$$

where K is a function of bounded variation.

- In the case of wavelet density estimators, the corresponding class of functions is

$$\mathcal{F}_\phi = \left\{ \sum_{k \in \mathbb{Z}} \phi(2^j y - k)\phi(2^j(\cdot) - k) : y \in \mathbb{R}, j \in \mathbb{N} \cup \{0\} \right\},$$

where ϕ is an α-Hölder continuous function with bounded support for some $\alpha \in (0, 1]$.

- Properties of convolution or of wavelet density estimators follow as a consequence of the fact that these classes are of VC type.

VC Hull

- **Definition 3.6.13** Given a class of functions $\mathcal{F}$, $co(\mathcal{F})$ is defined as the convex hull of $\mathcal{F}$,

$$co(\mathcal{F}) = \left\{ \sum_{f\in\mathcal{F}} \lambda_f f : f \in \mathcal{F}, \sum_f \lambda_f = 1, \lambda_f \geq 0, \lambda_f \neq 0 \text{ only for finitely many } f \right\},$$

and $\overline{co}(\mathcal{F})$ is defined as the pointwise sequential closure of $co(\mathcal{F})$. If $\mathcal{F}$ is VC subgraph, then $\overline{co}(\mathcal{F})$ is a VC hull class of functions.

Theorem 3.6.17

- **Theorem 3.6.17** Let Q be a probability measure on $(S, \mathcal{S})$ and let $\mathcal{F}$ be a collection of measurable functions with envelope $F \in L^2(Q)$ s.t.

$$N\Big(\mathcal{F}, L^2(Q), \epsilon \|F\|_{L^2(Q)}\Big) \le C\epsilon^{-w}, \quad 0 < \epsilon \le 1.$$

Then there exists a constant K depending only on C and w s.t.

$$\log N\Big(\overline{co}(\mathcal{F}), L^2(Q), \epsilon \|F\|_{L^2(Q)}\Big) \le K\epsilon^{-2w/(w+2)}, \quad 0 < \epsilon \le 1.$$

Example

- **Example 3.6.14** Let $\mathcal{F}$ be the class of monotone nondecreasing functions $f\colon \mathbb{R} \to [0,1]$. Then $\mathcal{F} \subseteq \overline{co}(\mathcal{G})$, where $\mathcal{G} = \{\mathbf{I}_{(x,\infty)}, \mathbf{I}_{[x,\infty)} : x \in \mathbb{R}\}$.
- Note that $N(\mathcal{G}, L^2(Q), \epsilon) \leq 2\epsilon^{-2}$ [Special case of Exercise 3.6.6].
- By Thm. 3.6.17, we have

$$\log N(\mathcal{F}, L^2(Q), \epsilon) \leq K/\epsilon, \quad 0 < \epsilon < 1,$$

for some $K < \infty$.

Proof of Theorem 3.6.17

- W.L.O.G, we assume $\mathcal{F}$ is finite.
- Set $u = \frac{1}{2} + \frac{1}{w}$ and $L = C^{1/w}\|F\|_{L^2(Q)}$.
- Since $N(\mathcal{F}, L^2(Q), \epsilon\|F\|_{L^2(Q)}) \leq C\epsilon^{-w}$, we have

$$N(\mathcal{F}, L^2(Q), Ln^{-1/w}) \leq n,$$

  and let $\mathcal{F}_n$ denote the collection of the centres of such a covering.
- It is enough to show that there exist constant C_k, D_k s.t. $\sup_{k\in\mathbb{N}} \max(C_k, D_k) < \infty$, and $q > 1$, satisfying

$$\log N(co(\mathcal{F}_{nk^q}), L^2(Q), C_k L n^{-u}) \leq D_k n, \quad n, k \geq 1. \tag{3.240}$$

- Note that given n, there exist $k < \infty$ s.t. $\mathcal{F}_{nk^q} = \mathcal{F}$.

Proof of Theorem 3.6.17

- **Lemma 3.6.16** Let $\mathcal{F}$ be a collection of measurable functions with envelope $F \in L^2(Q)$ s.t. $N(\mathcal{F}, L^2(Q), \epsilon\|F\|_{L^2(Q)}) \leq C\epsilon^{-w}, \quad 0 < \epsilon \leq 1.$ Set u, L as in a previous page. For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ be a maximal $Ln^{-1/w}$-separated subset of $\mathcal{F}$ for the $L^2(Q)$-norm. Then there exists $C_1 < \infty$ depending only on C and w s.t.

$$\log N\Big(co(\mathcal{F}_n), L^2(Q), C_1 L n^{-u}\Big) \leq n, \quad n \in \mathbb{N}.$$

- Using Lemma 3.6.16, (3.240) holds for $k = 1$ and all n with $C_1 < \infty$ and $D_1 = 1$.

Proof of Theorem 3.6.17

- For a induction, assume that (3.240) holds for $k-1$ and all n.
- And we have

$$co(\mathcal{F}_{nk^q}) \subseteq co(\mathcal{F}_{n(k-1)^q}) + co(\mathcal{G}_{n,k}),$$

  where $\mathcal{G}_{n,k}$ is a collection of at most nk^q functions of $L^2(Q)$-norm at most $L(n(k-1)^q)^{-1/w}$.
- Using Lemma 3.6.15 with $\epsilon = Lk^{-2}n^{-u}/(2L(n(k-1)^q)^{-1/2})$, (3.240) holds for k and for all n.
- **Lemma 3.6.15** Let $\mathcal{F} = \{f_1, \ldots, f_n\}$ be a collection of n functions in $L^2(Q)$. Then, for all $\epsilon > 0$,

$$N(co(\mathcal{F}), L^2(Q), \epsilon(\text{diam}\mathcal{F})) \leq (e + en\epsilon^2)^{2/\epsilon^2}.$$