MATH 281C: Mathematical Statistics

Lecture 10

1 VC Subgraph and Fat Shattering Dimensions

For Boolean function classes $\mathcal{F}$, we have seen that the VC dimension gives useful upper bounds on covering numbers

$$\sup_Q \log M(\epsilon, \mathcal{F}, L^2(Q)) \leq \left(\frac{c_1}{\epsilon}\right)^{c_2 \mathrm{VC}(\mathcal{F})} \quad \text{for all } 0 < \epsilon \leq 1. \tag{10.1}$$

What happens for function classes that are not Boolean? We will see today that there exist two notions of combinatorial dimension for general function classes, which allow control of covering numbers via bounds similar to (10.1). These are the notions of *VC subgraph dimension* and *fat shattering dimension* which we will go over today.

1.1 VC Subgraph Dimension

The VC subgraph dimension of $\mathcal{F}$ is simply the VC dimension of the Boolean class obtained by taking the indicators of the subgraphs of functions in $\mathcal{F}$. To formally define this, let us first define the notion of subgraph of a function.

Definition 1.1 (Subgraph)**.** For a function $f : \mathcal{X} \to \mathbb{R}$, its subgraph $\mathrm{sg}(f)$ is a subset of $\mathcal{X} \times \mathbb{R}$ defined by

$$\mathrm{sg}(f) = \{(x,t) \in \mathcal{X} \times \mathbb{R} : t < f(x)\}.$$

In other words, $\mathrm{sg}(f)$ consists of all points that lie below the graph of the function f.

We can now define the VC subgraph dimension of $\mathcal{F}$ as:

Definition 1.2 (VC subgraph dimension)**.** The VC subgraph dimension of $\mathcal{F}$ is defined as the VC dimension of the Boolean class $\{I_{\mathrm{sg}(f)} : f \in \mathcal{F}\}$. We denote it by $\mathrm{VC}(\mathcal{F})$.

The VC subgraph dimension can be related to covering numbers in the same way as (10.1). This is done in the following result.

Theorem 1.1. Suppose $\mathcal{F}$ is a class of functions with envelope F, and $\mathrm{VC}(\mathcal{F}) = D$. Then

$$\sup_Q M(\epsilon \|F\|_{L^2(Q)}, \mathcal{F}, L^2(Q)) \leq \left(\frac{c_1}{\epsilon}\right)^{c_2 D} \quad \text{for all } 0 < \epsilon \leq 1, \tag{10.2}$$

where $c_1, c_2 > 0$ are universal constants.

Proof. The idea is relate the $L^2(Q)$ norm between two functions in $\mathcal{F}$ to an L^2 norm between their subgraphs. Fix $f, g \in \mathcal{F}$, and write

$$\int |f-g|^2 \mathrm{d}Q \le \int 2F(x)|f(x)-g(x)|\mathrm{d}Q(x).$$

Note that, for every two real numbers a and b,

$$|a-b| = \int |I(t<a) - I(t<b)|\mathrm{d}t.$$

This gives

$$\begin{aligned}
\int |f-g|^2 \mathrm{d}Q &\le \int 2F(x)|f(x)-g(x)|\mathrm{d}Q(x) \\
&= \int 2F(x)\Big(\int |I\{t<f(x)\} - I\{t<g(x)\}|\,\mathrm{d}t\Big)\mathrm{d}Q(x) \\
&= \int\int |I_{\mathrm{sg}(f)}(x,t) - I_{\mathrm{sg}(g)}(x,t)| 2F(x)\,\mathrm{d}Q(x)\mathrm{d}t \\
&= \int\int_{(x,t):|t|\le F(x)} |I_{\mathrm{sg}(f)}(x,t) - I_{\mathrm{sg}(g)}(x,t)| 2F(x)\,\mathrm{d}Q(x)\mathrm{d}t \\
&= \Big\{\int\int_{(x,t):|t|\le F(x)} 2F(x)\mathrm{d}Q(x)\mathrm{d}t\Big\} \\
&\quad \times \int\int_{(x,t):|t|\le F(x)} |I_{\mathrm{sg}(f)}(x,t) - I_{\mathrm{sg}(g)}(x,t)| \frac{2F(x)\mathrm{d}Q(x)\mathrm{d}t}{\int\int_{(x,t):|t|\le F(x)} 2F(x)\mathrm{d}Q(x)\mathrm{d}t} \\
&= 4\int F^2(x)\mathrm{d}Q(x) \\
&\quad \times \int\int_{(x,t):|t|\le F(x)} |I_{\mathrm{sg}(f)}(x,t) - I_{\mathrm{sg}(g)}(x,t)|^2 \frac{2F(x)\mathrm{d}Q(x)\mathrm{d}t}{\int\int_{(x,t):|t|\le F(x)} 2F(x)\mathrm{d}Q(x)\mathrm{d}t}.
\end{aligned}$$

We have thus proved that

$$\|f-g\|_{L^2(Q)} \le 2\|F\|_{L^2(Q)}\|I_{\mathrm{sg}(f)} - I_{\mathrm{sg}(g)}\|_{L^2(Q')}, \tag{10.3}$$

where Q' is a probability measure on $\mathcal{X}\times\mathbb{R}$ whose density with respect to $Q\times$ Lebesgue is proportional to

$$2F(x)I\{(x,t) : |t| \le F(x)\}.$$

It then follows from (10.3) that

$$M\Big(\epsilon\|F\|_{L^2(Q)}, \mathcal{F}, L^2(Q)\Big) \le M\Big(\epsilon/2, \{I_{\mathrm{sg}(f)}, f\in\mathcal{F}\}, L^2(Q')\Big).$$

To bound the RHS, we simply use the earlier result for Boolean classes; see (10.1). This completes the proof of Theorem 1.1. □

The following is an immediate corollary of Theorem 1.1 and our main bound on the expected suprema of empirical processes.

Corollary 1.1. If $\mathcal{F}$ has envelop F and VC subgraph dimension D, then

$$\mathbb{E} \sup_{f \in \mathcal{F}} |P_n f - Pf| \le C \|F\|_{L^2(P)} \sqrt{\frac{D}{n}}. \tag{1.1}$$

Proof. Our main bound on the expected suprema of empirical processes gives

$$\mathbb{E} \sup_{f \in \mathcal{F}} |P_n f - Pf| \le C \|F\|_{L^2(P)} \frac{J(F, \mathcal{F})}{\sqrt{n}},$$

and we bound $J(F, \mathcal{F})$ (using Theorem 1.1) as

$$\begin{aligned} J(F, \mathcal{F}) &= \int_0^1 \sqrt{1 + \log \sup_Q M(\epsilon \|F\|_{L^2(Q)}, \mathcal{F}, L^2(Q))}\, \mathrm{d}\epsilon \\ &\le \int_0^1 \sqrt{1 + c_2 D \log(c_1/\epsilon)}\, \mathrm{d}\epsilon \le 1 + D^{1/2} \int_0^1 \sqrt{c_2 \log(c_1/\epsilon)}\, \mathrm{d}\epsilon \le C D^{1/2}, \end{aligned}$$

which completes the proof. □

Example 1.1. In the previous lecture, we claimed that

$$\mathbb{E} \sup_{\theta \in \mathbb{R}: |\theta - \theta_0| \le \delta} |(P_n - P)(m_\theta - m_{\theta_0})| \le C \sqrt{\frac{\delta}{n}},$$

where $m_\theta(x) := I(\theta - 1 \le x \le \theta + 1)$. We gave a partial proof of this result. To complete the proof, it suffices to show that the function class

$$\{ I_{[\theta-1, \theta+1]} - I_{[\theta_0 - 1, \theta_0 + 1]} : \theta \in \mathbb{R} \}$$

has finite VC subgraph dimension (≤ 3).

Let us now look at a reformulation of the VC subgraph dimension. This defines the dimension directly in terms of the class $\mathcal{F}$ without going to subgraphs. This reformulation will also make clear the connection to fat shattering dimension.

We say that a subset $\{x_1, \dots, x_n\}$ of $\mathcal{X}$ is subgraph-shattered by $\mathcal{F}$ if there exist real numbers $t_1, \dots, t_n$ such that for every subset $S \subseteq \{1, \dots, n\}$, there exists $f \in \mathcal{F}$ with

$$(x_s, t_s) \notin \mathrm{sg}(f) \;\text{ for }\; s \in S \quad \text{and} \quad (x_s, t_s) \in \mathrm{sg}(f) \;\text{ for } s \notin S. \tag{10.4}$$

Note that (10.4) is equivalent to

$$f(x_s) \le t_s \;\text{ for }\; s \in S \quad \text{and} \quad f(x_s) > t_s \;\text{ for } s \notin S. \tag{10.5}$$

In words, we say that $\{x_1, \dots, x_n\}$ of $\mathcal{X}$ is subgraph-shattered by $\mathcal{F}$ if there exist *levels* $t_1, \dots, t_n$ such that for every subset $S \subseteq \{1, \dots, n\}$, there exists a function f which goes under the level for each x_s, $s \in S$ and strictly over the level for x_s, $s \notin S$.

The VC subgraph dimension $\mathrm{VC}(\mathcal{F})$ is defined as the maximum cardinality of a finite subset of $\mathcal{X}$ that is subgraph-shattered by $\mathcal{F}$.

Let us now describe a potential problem with using the VC subgraph dimension to control covering numbers. Let $\mathcal{M}$ denote the class of all non-decreasing functions $f : \mathbb{R} \to [-1, 1]$, i.e., $\mathcal{M}$ consists of all non-decreasing functions on $\mathbb{R}$ that are bounded in magnitude by 1. It turns out that

$$\sup_Q M(\epsilon, \mathcal{M}, L^2(Q)) \leq \exp\left(\frac{C}{\epsilon}\right) \quad \text{for all } \epsilon > 0. \tag{10.6}$$

It is also easy to see that the VC subgraph dimension of $\mathcal{M}$ equals ∞, that is, for every $n \geq 1$, there exists a finite subset of $\mathbb{R}$ that is subgraph-shattered by $\mathcal{M}$. Therefore, the notion of VC subgraph dimension is not useful here and Theorem 1.1 does not give anything meaningful for this class $\mathcal{M}$. It is actually possible to prove (10.6) using the notion of fat shattering dimension.

1.2 Fat Shattering Dimension

The notion of fat shattering dimension is particularly useful to upper bound the packing number of the class of (i) uniformly bounded monotone functions and (ii) uniformly bound functions with finite total variation. The total variation of a function $f : \mathbb{R} \to \mathbb{R}$ is defined as

$$\|f\|_{\text{TV}} = \sup_{n \geq 1} \sup_{x_1 < x_2 < \cdots < x_n} \sum_{i=1}^{n-1} |f(x_i) - f(x_{i+1})|.$$

Mendelson and Vershynin (2003) proved that the L_2-covering numbers of every uniformly bounded class of functions are exponential in its shattering dimension. This extends Dudley's theorem on classes of $\{0, 1\}$-valued functions, for which the shattering dimension is the Vapnik-Chervonenkis dimension.

References

Mendelson, S. and Vershynin, R. (2003). Entropy and the combinatorial dimension. *Inventiones mathematicae* **152** 37–55.

Rudelson, M. and Vershynin, R. (2006). Combinatorics of random processes and sections of convex bodies. *Annals of Mathematics* **164** 603–648.