Stat210B: Theoretical Statistics	**Lecture Date: February 8, 2007**

Lecture 7

Lecturer: Michael I. Jordan *Scribe: Kurt Miller*

1 Properties of VC-classes

1.1 VC preservation

Let $\mathcal{C}$ and $\mathcal{D}$ be VC-classes (i.e. classes with finite VC-dimension). Then so are

- $\{C^{\complement} : C \in \mathcal{C}\}$
- $\{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$
- $\{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$
- $\phi(C)$ where ϕ is 1-1
- $\{C \times D : C \in \mathcal{C}, D \in \mathcal{D}\}$

1.2 Half spaces

Let $\mathcal{G}$ be a finite-dimensional vector space of functions. Let $\mathcal{C} = \{g \geq 0 : g \in \mathcal{G}\}$ or more formally $\mathcal{C} = \{\{\omega : g(\omega) \geq 0\} : g \in \mathcal{G}\}$. Then $V^{\mathcal{C}} \leq \dim \mathcal{G} + 1$.

1.3 Subgraphs

Definition 1. A *subgraph* of $f : \mathcal{X} \to \mathcal{R}$ is the subset $\mathcal{X} \times \mathcal{R}$ given by $\{(x, t) : t \leq f(x)\}$.

A collection $\mathcal{F}$ is a *VC-subgraph class* if the collection of subgraphs is a VC-class.

2 Covering Number

We now begin to explore a more powerful method of defining complexity than VC-dimension.

2.1 Definitions

Definition 2 (Covering Number). (Pollard, 1984, p. 25) Let Q be a probability measure on S and $\mathcal{F}$ be a class of functions in $\mathcal{L}^1(Q)$, i.e. $\forall f \in \mathcal{F}$, $\int |f| dQ < \infty$. For each $\varepsilon > 0$ define the $\mathcal{L}_1$ covering number $N_1(\varepsilon, Q, \mathcal{F})$ as the smallest value of m for which there exist functions $g_1, \ldots, g_m$ (not necessarily in $\mathcal{F}$) such that $\min_j Q|f - g_j| \leq \varepsilon$ for each f in $\mathcal{F}$. For definiteness set $N_1(\varepsilon, Q, \mathcal{F}) = \infty$ if no such m exists.

Note that the set $\{g_j\}$ that achieves this minimum is not necessarily unique.

Definition 3 (Metric Entropy). Define $H_1(\varepsilon, Q, \mathcal{F}) = \log N_1(\varepsilon, Q, \mathcal{F})$ as the $\mathcal{L}_1$ metric entropy of $\mathcal{F}$.

More generally, $H_p(\varepsilon, Q, \mathcal{F})$ uses the $\mathcal{L}_p(Q)$ norm. Write this as $\|g\|_{p,Q} = (\int |g|^p dQ)^{1/p}$.

Definition 4 (Totally bounded). A class is called *totally bounded* if $\forall \varepsilon,\ H_p(\varepsilon, Q, \mathcal{F}) < \infty$

Another kind of entropy:

Definition 5 (Entropy with bracketing). Let $N_{p,B}(\varepsilon, Q, \mathcal{F})$ be the smallest value of m for which there exist pairs of functions $\{(g_j^L, g_j^U)\}_{j=1}^m$ such that $\forall j$, $\|g_j^U - g_j^L\|_{p,Q} < \varepsilon$ and $\forall f \in \mathcal{F}, \exists j(f)$ s.t. $g_{j(f)}^L \leq f \leq g_{j(f)}^U$. Then we define the *entropy with bracketing* as $H_{p,Q}(\varepsilon, Q, \mathcal{F}) = \log N_{p,Q}(\varepsilon, Q, \mathcal{F})$.

Finally, using $\|g\|_\infty \triangleq \sup_{x \in \mathcal{X}} |g(x)|$, let $N_\infty(\varepsilon, \mathcal{F})$ be the smallest m such that there exists a set $\{g_j\}_{j=1}^m$ such that $\sup_{f \in \mathcal{F}} \min_{j=1,\ldots,m} \|f - g_j\|_\infty < \varepsilon$. Then $H_\infty(\varepsilon, \mathcal{F}) = \log N_\infty(\varepsilon, \mathcal{F})$.

2.2 Relationship of the various entropies

Using the definitions above, we have that

1. $H_1(\varepsilon, Q, \mathcal{F}) \leq H_{p,B}(\varepsilon, Q, \mathcal{F})$, $\forall \varepsilon > 0$
2. $H_{p,B}(\varepsilon, Q, \mathcal{F}) \leq H_\infty(\varepsilon/2, \mathcal{F})$, $\forall \varepsilon > 0$

Can these quantities be computed for normal classes of functions? Yes, but you would generally look them up in a big book. We'll look at how to compute one of these quantities here.

2.3 Examples

Example 6. Let $\mathcal{F} = \{f : [0,1] \to [0,1],\ |f'| \leq 1\}$ (i.e. functions from $[0,1]$ to $[0,1]$ with first derivatives bounded by 1). Then $H_\infty(\varepsilon, \mathcal{F}) \leq A\frac{1}{\varepsilon}$ where A is a constant that we will compute.

Proof. Let $0 = a_0 < a_1 < \cdots < a_m = 1$ where $a_k = k\varepsilon$ and $k = 0, \ldots, m$. Let $B_1 = [a_0, a_1]$ and $B_k = (a_{k-1}, a_k]$. For each $f \in \mathcal{F}$, define

$$\tilde{f} = \sum_{k=1}^{m} \varepsilon \left\lfloor \frac{f(a_k)}{\varepsilon} \right\rfloor 1_{B_k}$$

$\tilde{f}$ takes on values in εk where k is an integer. We also have $\|\tilde{f} - f\|_\infty \leq 2\varepsilon$, because $|\tilde{f}(a_{k-1}) - f(a_{k-1})| \leq \varepsilon$ by construction and $|f(a_k)^- - f(a_{k-1})| \leq \varepsilon$ since f' is bounded by 1.

We now count the number of possible $\tilde{f}$ obtained by this construction. At a_0, there are $\lfloor 1/\varepsilon \rfloor + 1$ choices for $\tilde{f}(a_0)$ since $\tilde{f}$ only takes on values of εk in $[0,1]$. Furthermore, combining previous results gives us

$$\begin{aligned} |\tilde{f}(a_k) - \tilde{f}(a_{k-1})| &\leq |\tilde{f}(a_k) - f(a_k)| + |f(a_k) - f(a_{k-1})| + |f(a_{k-1}) - \tilde{f}(a_{k-1})| \\ &\leq 3\varepsilon. \end{aligned}$$

Therefore, having chosen $\tilde{f}(a_{k-1})$, $\tilde{f}$ can take on at most 7 distinct values at a_k. Therefore

$$N_\infty(2\varepsilon, \mathcal{F}) \leq \left(\left\lfloor \frac{1}{\varepsilon} \right\rfloor + 1 \right) 7^{\lfloor 1/\varepsilon \rfloor}$$

which gives us that

$$H_\infty(2\varepsilon, \mathcal{F}) \leq \frac{1}{\varepsilon} \log 7 + \log(\lfloor 1/\varepsilon \rfloor + 1)$$

so our constant can be chosen as any constant that $> \log 7$. □

A seminal paper in this field is by Birman and Solomjak in 1967. They present other examples of metric entropy calculations, including:

Example 7. Let $\mathcal{F} = \{f : [0,1] \to [0,1] : \int (f^{(m)}(x))^2 dx \leq 1\}$. Then $H_\infty(\varepsilon, \mathcal{F}) \leq A\varepsilon^{-1/m}$.

Example 8. Let $\mathcal{F} = \{f : \mathcal{R} \to (0,1) : f \text{ is increasing}\}$. Then $H_{p,B}(\varepsilon, Q, \mathcal{F}) \leq A\frac{1}{\varepsilon}$.

Example 9. Let $\mathcal{F} = \{f : \mathcal{R} \to [0,1] : \int |df| \leq 1\}$, the class of bounded variation. Then $H_{p,B}(\varepsilon, Q, \mathcal{F}) \leq A\frac{1}{\varepsilon}$.

Lemma 10 (Ball covering lemma). *A ball $B_d(R)$ in $\mathcal{R}^d$ of radius R can be covered by*

$$\left(\frac{4R + \varepsilon}{\varepsilon} \right)^d$$

balls of radius ε.

Proof. Let $\{c_j\}_{j=1}^m$ be a packing of size ε (Euclidean norm). This implies that balls of radius ε with centers at $\{c_j\}$ cover $B_d(R)$ (otherwise we could add more points c_j to the packing). Let B_j be the ball of radius $\varepsilon/4$ centered at c_j. We must have that $B_i \cap B_j$ is empty for $i \neq j$. Therefore $\{B_j\}$ are disjoint and

$$\cup_j B_j \subset B_d(R + \varepsilon/4).$$

A ball of radius ρ has volume $C_d \rho^d$ where C_d is a constant that depends on the dimension d. Therefore, the volume of the union $\cup_j B_j$ is $MC_d(\varepsilon/4)^d$ and since it is a subset of $B_d(R + \varepsilon/4)$, we have

$$MC_d \left(\frac{\varepsilon}{4} \right)^d \leq C_d \left(R + \frac{\varepsilon}{4} \right)^d.$$

With a simple manipulation of this equation, we get that

$$M \leq \left(\frac{4R + \varepsilon}{\varepsilon} \right)^d$$

□

References

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.