CMSC 35900 (Spring 2008) Learning Theory **Lecture: 16**

Fat Shattering Dimension and Covering Numbers

Instructors: Sham Kakade and Ambuj Tewari

In this lecture, we will prove a result due to Alon, Ben-David, Cesa-Bianchi and Haussler that bound the covering number of a class in terms of its fat shattering dimension. This provides a result analogous to Sauer's lemma. As you remember, Sauer's lemma gave us a bound on the growth function in terms of its VC dimension.

1 Functions with Finite Range

Before we prove the result we need a few definitions. Suppose $\mathcal{X}$ is some set and let $B = \{0, 1, \ldots, b\}$. Let $\mathcal{F} \subseteq B^{\mathcal{X}}$ be a class of B-valued functions on $\mathcal{X}$. Two functions $f, g \in \mathcal{F}$ are said to be *separated* if

$$\exists x \in \mathcal{X} \text{ s.t. } |f(x) - g(x)| \geq 2 \ .$$

That is, they are 2-separated in the ℓ^∞ metric where

$$\ell^\infty(f, g) := \max_{x \in \mathcal{X}} |f(x) - g(x)| \ .$$

A class $\mathcal{F}$ is said to be pairwise separated iff f, g are pairwise separated for all $f, g \in \mathcal{F}$.

Definition 1.1. *Let $\mathcal{F} \subseteq B^{\mathcal{X}}$. We say that $\mathcal{F}$* strongly shatters *$X = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ if there exists $\mathbf{s} = (s_1, \ldots, s_n) \in B^n$ such that for all $E \subseteq \{x_1, \ldots, x_n\}$, there exists $f_E \in \mathcal{F}$ such that*

$$\begin{aligned} &\forall x_i \in E, && f_E(x_i) \geq s_i + 1 \\ &\forall x_i \in X - E, && f_E(x_i) \leq s_i - 1 \end{aligned}$$

In this case we also say that $\mathcal{F}$ strongly shatters X according to $\mathbf{s}$. *The strong dimension of $\mathcal{F}$, denoted by* $\mathrm{Sdim}(\mathcal{F})$*, is the size of a largest strongly shattered set.*

We will shift our attention from real valued functions to ones taking values in a finite set by a simple discretization.

Definition 1.2. *Let $f : \mathcal{X} \to [0, 1]$ be a function. For $\alpha > 0$, define its discretization f^α as,*

$$f^\alpha(x) := \left\lfloor \frac{f(x)}{\alpha} \right\rfloor \ .$$

If $\mathcal{F}$ is a function class, define

$$\mathcal{F}^\alpha := \{f^\alpha \,|\, f \in \mathcal{F}\} \ .$$

Note that f^α takes value in the set $\{0, \ldots, \lfloor 1/\alpha \rfloor\}$. The following lemma relates the combinatorial dimensions and packing numbers of the classes $\mathcal{F}$ and $\mathcal{F}^\alpha$.

Recall that we defined the covering number $\mathcal{N}_\infty(\alpha, \mathcal{F}, x_{1:n})$ in an earlier lecture. We define the corresponding packing number as

$$\mathcal{M}_\infty(\alpha, \mathcal{F}, x_{1:n}) := \mathcal{M}_{\ell^\infty_{x_{1:n}}}(\alpha, \mathcal{F}) \ ,$$

where

$$l^\infty_{x_{1:n}}(f, g) = \max_{i \in [n]} |f(x_i) - g(x_i)| \ .$$

Lemma 1.3. *Let $\mathcal{F} \subseteq [0,1]^{\mathcal{X}}$ and $\alpha > 0$. We have*

1. $\mathrm{Sdim}(\mathcal{F}^\alpha) \le \mathrm{fat}_{\alpha/2}(\mathcal{F})$
2. *For any $x_{1:n}$, $\mathcal{M}_\infty(\alpha, \mathcal{F}, x_{1:n}) \le \mathcal{M}_\infty(2, \mathcal{F}^{\alpha/2}, x_{1:n})$*

To prove a result bounding the ∞-covering number in terms of the fat shattering dimension, we need the following combinatorial lemma whose proof we will give shortly.

Lemma 1.4. *Let $\mathcal{X}$ be a finite set with $|\mathcal{X}| = n$ and $B = \{0, 1, \ldots, b\}$. Let $\mathcal{F} \subseteq B^{\mathcal{X}}$ be such that $\mathrm{Sdim}(\mathcal{F}) = d$. Then we have,*

$$\mathcal{M}_{\ell^\infty}(2, \mathcal{F}) < 2(n(b+1)^2)^{\lceil \log y \rceil} ,$$

where $y = \sum_{i=1}^{d} \binom{n}{i} b^i$.

Using the above lemma, we can prove a result relating covering numbers to fat shattering dimension.

Theorem 1.5. *Let $\mathcal{F} \subseteq [0,1]^{\mathcal{X}}$ and $\alpha \in [0,1]$. Suppose $d = \mathrm{fat}_{\alpha/4}(\mathcal{F})$. Then*

$$\mathcal{N}_\infty(\alpha, \mathcal{F}, n) < 2\left(n\left(\frac{2}{\alpha}+1\right)^2\right)^{\left\lceil d \log\left(\frac{2en}{d\alpha}\right)\right\rceil} .$$

Proof. Using the fact that covering numbers are bounded by packing numbers, Lemma 1.3, part 2 and Lemma 1.4, we get

$$\begin{aligned} \mathcal{N}_\infty(\alpha, \mathcal{F}, n) &= \sup_{x_{1:n}} \mathcal{N}_\infty(\alpha, \mathcal{F}, x_{1:n}) \\ &\le \sup_{x_{1:n}} \mathcal{M}_\infty(\alpha, \mathcal{F}, x_{1:n}) \\ &\le \sup_{x_{1:n}} \mathcal{M}_\infty(2, \mathcal{F}^{\alpha/2}, x_{1:n}) \\ &< 2(n(b+1)^2)^{\lceil \log y \rceil} , \end{aligned}$$

where $b = \lfloor 2/\alpha \rfloor$ and $y = \sum_{i=1}^{d'} \binom{n}{i} b^i$ with $d' = \mathrm{Sdim}(\mathcal{F}^{\alpha/2})$. By Lemma 1.3, part 1, $d' \le \mathrm{fat}_{\alpha/4}(\mathcal{F}) = d$. Therefore,

$$\begin{aligned} y &\le \sum_{i=1}^{d} \binom{n}{i} b^i \\ &\le b^d \sum_{i=1}^{d} \binom{n}{i} \le b^d \left(\frac{en}{d}\right)^d . \end{aligned}$$

Thus, $\log y \le d \log(ben/d)$. □

The rest of this lecture is devoted to proving Lemma 1.4.

Proof of Lemma 1.4. Fix $b \ge 2$ as the result trivially holds otherwise. For $h \ge 2$, $n \ge 1$, define the function

$$\begin{aligned} t(h, n) = \max\{k \mid \forall F \subseteq \mathcal{F}, |F| = h, F \text{ pairwise separated} \\ \Rightarrow F \text{ strongly shatters at least } k\ (X, \mathbf{s}) \text{ pairs}\} . \end{aligned}$$

When we say F strongly shatters a pair $(X, \mathbf{s})$, we mean F strongly shatters X according to $\mathbf{s}$. Note that $t(h, n) = \infty$ when no pairwise separated F of cardinality h exists. Because of the following claim, it suffices to show

$$t\left(2(n(b+1)^2)^{\lceil \log y \rceil}, n\right) \ge y . \tag{1}$$

Claim 1.6. *If $t(h,n) \geq y$ for some h and $\mathrm{Sdim}(\mathcal{F}) \leq d$ then*

$$\mathcal{M}_{\ell^\infty}(2,\mathcal{F}) < h \ .$$

Proof. For the sake of deriving a contradiction, suppose $\mathcal{M}_{\ell^\infty}(2,\mathcal{F}) \geq h$. This means there is a pairwise separated set F of cardinality at least h. Since $t(h,n) \geq y$, F strongly shatters at least y $(X,\mathbf{s})$ pairs. On the other hand, since $\mathrm{Sdim}(\mathcal{F}) \leq d$, if F strongly shatters $(X,\mathbf{s})$ then $|X| \leq d$. For any choice of X of size i (there are $\binom{n}{i}$ such choices), there are strictly less than b^i choices for $\mathbf{s}$. This is because if

$$(X, \mathbf{s} = (s_1,\ldots,s_{|X|}))$$

is strongly shattered then s_i's cannot be 0 or b. Thus, F can strongly shatter strictly less than

$$\sum_{i=1}^{d} \binom{n}{i} b^i = y$$

$(X,\mathbf{s})$ pairs. This gives us a contradiction. □

To prove (1) by induction, we will establish the following two claims,

$$t(2,n) \geq 1 \qquad n \geq 1 \ , \tag{2}$$
$$t(2mn(b+1)^2, n) \geq 2t(2m, n-1) \qquad m \geq 1, n \geq 2 \ . \tag{3}$$

Any separated functions f,g strongly shatters at least some singeton $X=\{x\}$ (choose any x such that $|f(x)-g(x)| \geq 2$), so $t(2,n) \geq 1$. To prove (3), consider a set F of $2mn(b+1)^2$ pairwise separated functions. If such a set does not exist then $t(2mn(b+1)^2,n) = \infty$ so (3) anyway holds. Pair up the functions in F arbitrarily to form $mn(b+1)^2$ pairs $\{f,g\}$. Call the set of these pairs P. For each pair f,g, fix an x on which they differ by at least 2 and denote it by $\chi(f,g)$. For $x \in \mathcal{X}$ and $i,j \in B$, $j > i+1$, define

$$\mathrm{bin}(x,i,j) = \{\{f,g\} \in P \,|\, \chi(f,g) = x,\ \{f(x),g(x)\} = \{i,j\}\} \ .$$

The number of bins is no more than $n\binom{b+1}{2} < n(b+1)^2/2$ and the numbers of pairs is $mn(b+1)^2$, so for some $x^* \in \mathcal{X}$, $i^*, j^* \in B$ such that $j^* > i^*+1$, we have

$$|\,\mathrm{bin}(x^*,i^*,j^*)| \geq 2m \ .$$

Now define the following two set of functions,

$$F_1 := \left\{ f \in \bigcup \mathrm{bin}(x^*,i^*,j^*) \,\middle|\, f(x^*) = i^* \right\} \ ,$$
$$F_2 := \left\{ f \in \bigcup \mathrm{bin}(x^*,i^*,j^*) \,\middle|\, f(x^*) = j^* \right\} \ .$$

Clearly $|F_1| = |F_2| = 2m$. The first important observation is that F_1 is pairwise separate on the domain $\mathcal{X} - \{x^*\}$ (which has cardinality $n-1$). This is because all $f \in F_1$ take value i^* on x^*. Similarly F_2 is pairwise separate on $\mathcal{X} - \{x^*\}$. Therefore, there exists sets U,V consisting of pairs $(X,\mathbf{s})$ such that F_1, F_2 strongly shatter pairs in U, V respectively. Further, $|U| \geq t(2m,n-1)$ and $|V| \geq t(2m,n-1)$. Any pair in $U \cup V$ is obviously shattered by F. Now consider any pair $(X,\mathbf{s}) \in U \cap V$. Then, $(x^* \cup X, (\lfloor \frac{i^*+j^*}{2} \rfloor, \mathbf{s}))$ is also shattered by F (remember that $j^* > i^*+1$). Thus, F strongly shatters

$$|U \cup V| + |U \cap V| = |U| + |V| \geq 2t(2m,n-1)$$

pairs. This completes the proof of (3).

Once we have (2) and (3), it easily follows that for $n > r \geq 1$,

$$t(2(n(b+1)^2)^r, n) \geq 2^r t(2, n-r) \geq 2^r \ .$$

Thus if $\lceil \log y \rceil < n$, we can set $r = \lceil \log y \rceil$ above and (1) follows. On the other hand, if $\lceil \log y \rceil \geq n$, then

$$2(n(b+1)^2)^{\lceil \log y \rceil} > (b+1)^n$$

which exceeds the total number of B-valued functions defined on a set of size n. Thus, a pairwise separated set F of size $2(n(b+1)^2)^{\lceil \log y \rceil}$ does not exist and hence

$$t(2(n(b+1)^2)^{\lceil \log y \rceil}, n) = \infty \ .$$

So (1) still holds.

$\square$