CMPUT 654 Fa 23: Theoretical Foundations of Machine Learning **Fall 2023**

Lecture 14: Covering Number Estimates (October 19)

Lecturer: Csaba Szepesvári *Scribes: Kushagra Chandak*

(Part 1). Let $\mathcal{G} \subseteq \{0,1\}^{\mathcal{Z}}$ be a **VC class**, that is, $\mathrm{VC}(\mathcal{G}) = d$. Then Sauer's lemma gives us $\mathcal{N}_\infty(\varepsilon, \mathcal{G}, n) \le O(d \ln(n/d))$. Using Haussler's result, we can bound empirical metric entropy without $\ln(n)$ dependence but $\ln(1/\varepsilon)$ dependence:

$$\ln \mathcal{M}(\varepsilon, \mathcal{G}, L_1(z_{1:n}) \le 1 + \ln(d+1) + \ln \frac{2e}{\varepsilon} \,.$$

We will see later how we can remove the $\ln n$ factor from uniform deviation bounds using generalized Haussler's result and a technique called chaining.

(Part 2). Recall that VC dimension was a combinatorial quantity, that is, it was defined for $\{0,1\}$-valued function classes. We next extend this notion to real-valued function classes.

Definition 14.1 (Subgraph function)**.** Let $\mathcal{G} : \mathcal{Z} \to \mathbb{R}$. The subgraph function for $g \in \mathcal{G}$ is a map $(g) : \mathcal{Z} \times \mathbb{R} \to \{0,1\}$ given by $(z,t) \mapsto \mathbb{I}(t \le g(z))$.

Definition 14.2 (VC-subgaph dimension)**.** The VC-subgraph dimension (also called Pollard pseudo dimension) is defined as $\mathrm{VC}(\mathcal{G}) = \mathrm{VC}((\mathcal{G})$.

Proposition 14.3. *Let $\mathcal{Z} = \mathbb{R}^d$ and $w \in \mathbb{R}^d$. Let $f_w : \mathcal{Z} \to \mathbb{R}$ given by $z \mapsto w^\top z$ and $\mathrm{LIN}_d = \left\{ f_w : w \in \mathbb{R}^d \right\}$. Then $\mathrm{VC}(\mathrm{LIN}_d) \le d+1$.*

It is worthwhile to remember that VC-dimension is monotone for inclusion. That is if $A \subseteq B$, then $\mathrm{VC}(A) \le \mathrm{VC}(B)$.

Proposition 14.4. *Let $h : \mathbb{R} \to \mathbb{R}$ be a monotone function. Then $\mathrm{VC}_{(}h \circ \mathcal{G}) \le \mathrm{VC}_{(}\mathcal{G})$.*

Theorem 14.5. *Let $\mathcal{G} \subseteq [0,1]^{\mathcal{Z}}$ and $d = \mathrm{VC}_{(}\mathcal{G})$. Then*

1. *For $1 \le p < \infty$*

$$\begin{aligned} \mathcal{M}(\varepsilon, \mathcal{G}, L_p(\mu)) &\le \mathcal{M}(\varepsilon, (\mathcal{G}), L_p(\mu \otimes U[0,1]) \\ &\le 1 + \ln(d+1) + \ln \frac{2e}{\varepsilon^p} \,. \end{aligned}$$

2. *For $0 \le \varepsilon < 1$*

$$\ln N_\infty(\varepsilon, \mathcal{G}, n) \le d \ln \max\left(2, \frac{en(1+\varepsilon)}{\varepsilon}\right) \,.$$

14.1 Packing and Covering for Bounded Function Classes

A function class that has a uniform deviation bound with $d \log \frac{c}{\varepsilon}$ for some c is sometimes called a parametric class whereas classes with $\frac{1}{\varepsilon}$ rates are called non-parametric classes.