

Lecture 12: VC Dimension (October 12)

Lecturer: Csaba Szepesvári

Scribes: Kushagra Chandak

Note: *LaTeX* template courtesy of UC Berkeley EECS dept. ([link to directory](#))

Disclaimer: These notes have **not** been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.

To get uniform deviation bounds and oracle inequalities, we need empirical covering numbers. The empirical L_p norm for a function g is defined as

$$\|g\|_{L_p(z_{1:n})}^p = \frac{1}{n} \sum_{i=1}^n |g(z_i)|^p.$$

If $p > q$, then $\mathcal{N}_p > \mathcal{N}_q$.

Definition 12.1 (L_∞ empirical cover). Fix $\varepsilon > 0$. Let $\mathcal{G} \subseteq \{0, 1\}^{\mathcal{Z}}$ and $z_{1:n} \in \mathcal{Z}^n$. The set of functions $g_1, \dots, g_m \in \mathcal{G}$ is (ε, L_∞) -empirical cover of \mathcal{G} if for every $g \in \mathcal{G}$ there exists $j \in [m]$ such that $d_{z_{1:n}}^\infty(g, g_j) \leq \varepsilon$, where

$$d_{z_{1:n}}^\infty(g, g_j) = \max_{1 \leq i \leq n} |g(z_i) - g_j(z_i)|.$$

The L_∞ covering number is $\mathcal{N}_\varepsilon = \mathcal{N}_\infty(\varepsilon, \mathcal{G}, z_{1:n}) = \min \{m : (\varepsilon, L_\infty)\text{-cover exists}\}$.

Now consider $\varepsilon < 1$. For $g, g' \in \mathcal{G}$ if $d_{z_{1:n}}^\infty(g, g') < 1$, then $g(z_i) = g'(z_i)$ for all $i \in [n]$.

Denote by $\mathcal{G}(z_{1:n}) = \{g(z_{1:n}) : g \in \mathcal{G}\}$ the total number of behaviors of \mathcal{G} on $z_{1:n}$. It is the total number of binary vectors of length n when g is evaluated on $z_{1:n}$ for all $g \in \mathcal{G}$. The maximum possible behaviors is 2^n .

add context...

$$\begin{aligned} \mathcal{N}_\infty(\varepsilon, \mathcal{G}, n) &= \sup_{z_{1:n} \in \mathcal{Z}^n} \mathcal{N}_\infty(\varepsilon, \mathcal{G}, z_{1:n}) \\ &= \sup_{z_{1:n} \in \mathcal{Z}^n} |\mathcal{G}(z_{1:n})|. \end{aligned}$$

Example: Let $\mathcal{G} = \{g_{z_0} : z_0 \in \mathcal{Z}\}$ with $g_{z_0}(z) = \mathbb{I}(z = z_0)$. The total number of behaviors in this case is $n + 1$. So, $\mathcal{N}_\infty(\varepsilon, \mathcal{G}, n) = n + 1$.

VC-dimension tells us about

- characterization of L_∞ empirical covers;
- learnability with classification loss;
- PAC learnability is equivalent to $VC < \infty$.