VC-subgraph classes of functions

Let $\mathcal{F} = \{f : \mathcal{X} \mapsto \mathbb{R}\}$ and

$$C_f = \{(x,t) \in \mathcal{X} \times \mathbb{R} : 0 \le t \le f(x) \text{ or } f(x) \le t \le 0\}.$$

Define class of sets $\mathcal{C} = \{C_f : f \in \mathcal{F}\}$.

Definition 12.1. *If $\mathcal{C}$ is a VC class of sets, then $\mathcal{F}$ is VC-subgraph class of functions and, by definition, $VC(\mathcal{F}) = VC(\mathcal{C})$.*

Note that equivalent definition of C_f is

$$C'_f = \{(x,t) \in \mathcal{X} \times \mathbb{R} : f(x) \ge t\}.$$

Example 1. $\mathcal{C} = \{C \subseteq \mathcal{X}\}$, $\mathcal{F}(\mathcal{C}) = \{I(X \in C) : C \in \mathcal{C}\}$. *Then $\mathcal{F}(\mathcal{C})$ is VC-subgraph class if and only if $\mathcal{C}$ is a VC class of sets.*

Example 2. *Assume d functions are fixed: $\{f_1, \ldots, f_d\} : \mathcal{X} \mapsto \mathbb{R}$. Let*

$$\mathcal{F} = \left\{ \sum_{i=1}^{d} \alpha_i f_i(x) : \alpha_1, \ldots, \alpha_d \in \mathbb{R} \right\}.$$

Then $VC(\mathcal{F}) \le d + 1$. To prove this, it's easier to use the second definition.

Packing and covering numbers

Let $f, g \in \mathcal{F}$ and assume we have a distance function $d(f,g)$.

Example 3. *If $X_1, \ldots, X_n$ are data points, then*

$$d_1(f,g) = \frac{1}{n} \sum_{i=1}^{n} |f(X_i) - g(X_i)|$$

and

$$d_2(f,g) = \left(\frac{1}{n} \sum_{i=1}^{n} (f(X_i) - g(X_i))^2 \right)^{1/2}.$$

Definition 12.2. *Given $\varepsilon > 0$ and $f_1, \ldots, f_N \in \mathcal{F}$, we say that $f_1, \ldots, f_N$ are ε-separated if $d(f_i, f_j) > \varepsilon$ for any $i \ne j$.*

Definition 12.3. *The ε-packing number, $\mathcal{D}(\mathcal{F}, \varepsilon, d)$, is the maximal cardinality of an ε-separated set.*

Note that $\mathcal{D}(\mathcal{F}, \varepsilon, d)$ is decreasing in ε.

Definition 12.4. *Given $\varepsilon > 0$ and $f_1, \ldots, f_N \in \mathcal{F}$, we say that the set $f_1, \ldots, f_N$ is an ε-cover of $\mathcal{F}$ if for any $f \in \mathcal{F}$, there exists $1 \le i \le N$ such that $d(f, f_i) \le \varepsilon$.*

Definition 12.5. *The ε-covering number, $\mathcal{N}(\mathcal{F}, \varepsilon, d)$, is the minimal cardinality of an ε-cover of $\mathcal{F}$.*

Lemma 12.1.

$$\mathcal{D}(\mathcal{F}, 2\varepsilon, d) \le \mathcal{N}(\mathcal{F}, \varepsilon, d) \le \mathcal{D}(\mathcal{F}, \varepsilon, d).$$

Proof. To prove the first inequality, assume that $\mathcal{D}(\mathcal{F}, 2\varepsilon, d) > \mathcal{N}(\mathcal{F}, \varepsilon, d)$. Let the packing corresponding to the packing number $\mathcal{D}(\mathcal{F}, 2\varepsilon, d) = D$ be $f_1, \ldots, f_D$. Let the covering corresponding to the covering number $\mathcal{N}(\mathcal{F}, \varepsilon, d) = N$ be $f'_1, \ldots, f'_N$. Since $D > N$, there exist f_i and f_j such that for some f'_k

$$d(f_i, f'_k) \le \varepsilon \text{ and } d(f_j, f'_k) \le \varepsilon.$$

Therefore, by triangle inequality, $d(f_i, f_j) \le 2\varepsilon$, which is a contradiction.

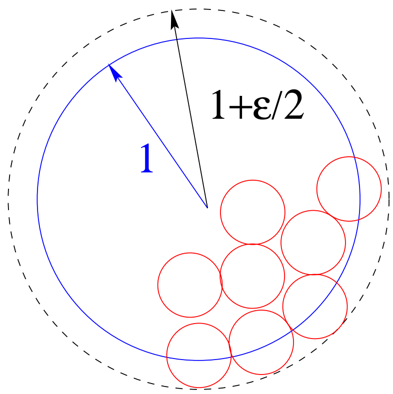
To prove the second inequality, assume $f_1, \ldots, f_D$ is an optimal packing. For any $f \in \mathcal{F}$, $f_1, \ldots, f_D, f$ would also be ε-packing if $d(f, f_i) > \varepsilon$ for all i. Since $f_1, \ldots, f_D$ is optimal, this cannot be true, and, therefore, for any $f \in \mathcal{F}$ there exists f_i such that $d(f, f_i) \le \varepsilon$. Hence $f_1, \ldots, f_D$ is also a cover. Hence, $\mathcal{N}(\mathcal{F}, \varepsilon, d) \le \mathcal{D}(\mathcal{F}, \varepsilon, d)$. □

Example 4. *Consider the L_1-ball $\{x \in \mathbb{R}^d, |x| \le 1\} = B_1(0)$ and $d(x, y) = |x - y|_1$. Then*

$$\mathcal{D}(B_1(0), \varepsilon, d) \le \left(\frac{2 + \varepsilon}{\varepsilon}\right)^d \le \left(\frac{3}{\varepsilon}\right)^d,$$

where $\varepsilon \le 1$. Indeed, let $f_1, \ldots, f_D$ be optimal ε-packing. Then the volume of the ball with $\varepsilon/2$-fattening (so that the center of small balls fall within the boundary) is

$$Vol\left(1 + \frac{\varepsilon}{2}\right) = C_d \left(1 + \frac{\varepsilon}{2}\right)^d.$$



Moreover, the volume of each of the small balls

$$Vol\left(\frac{\varepsilon}{2}\right) = C_d\left(\frac{\varepsilon}{2}\right)^d$$

and the volume of all the small balls is

$$DC_d\left(\frac{\varepsilon}{2}\right)^d.$$

Therefore,

$$D \leq \left(\frac{2+\varepsilon}{\varepsilon}\right)^d.$$

Definition 12.6. $\log \mathcal{N}(\mathcal{F}, \varepsilon, d)$ *is called metric entropy.*

For example, $\log \mathcal{N}(B_1(0), \varepsilon, d) \leq d \log \frac{3}{\varepsilon}$.