GENERALIZATION VIA COVERING NUMBERS

DEF (Cover) Let $(M, \|\cdot\|)$ be a metric space. Given $\mathcal{U} \subset M$, scale $\varepsilon > 0$, $V \subseteq \mathcal{U}$ is a (proper) cover when

$$\sup_{u \in \mathcal{U}} \inf_{v \in V} \|u - v\| \leq \varepsilon.$$

DEF (Improper cover) Given a set $\mathcal{U} \subset M$ and scale $\varepsilon > 0$, $V \subset M$ is an improper cover of $\mathcal{U}$ if

$$\sup_{u \in \mathcal{U}} \inf_{v \in V} \|u - v\| \leq \varepsilon.$$

DEF (Covering number)

$$\mathcal{N}(\mathcal{U}, \varepsilon, \|\cdot\|) = \inf \left\{ |V| \ : \ V \subseteq \mathcal{U}, \ \sup_{u \in \mathcal{U}} \inf_{v \in V} \|u - v\| \leq \varepsilon \right\}.$$

First Idea : ε-net argument

Given a class $\mathcal{F}$ of hypotheses, suppose that $\exists \{f_1, \dots, f_N\}$ where $N = \mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|)$ s.t. for any $f \in \mathcal{F}$, $\exists i \in \{1, \dots, N\}$ s.t. $\|f - f_i\| \leq \varepsilon$.

Informally if $\mathcal{F}$ is included in an $\|\cdot\|_\infty$-ball of radius c, it can be easily covered by $\left(\frac{c}{\varepsilon}\right)^d$ hypercubes of edge-length 2ε.

$$\Rightarrow \log \mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty) = O\left(d \log \frac{1}{\varepsilon}\right).$$

THEOREM 1 Suppose that $\ell(y, \hat{y}) \in [0, a]$ for all $y, \hat{y} \in \mathcal{Y}$, and $y \mapsto \ell(\hat{y}, y)$ is L-Lipschitz for all $\hat{y} \in \mathcal{Y}$. Given $\mathcal{F}$, $\{f_1, \dots, f_N\}$ be a minimal ε-cover of $\mathcal{F}$ w.r.t. $\|\cdot\|_\infty$. Then,

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\hat{R}_S(f) - R(f)| \leq 2L\varepsilon + 2a \cdot \sqrt{\frac{2 \log(2\mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty))}{n}}$$

Proof For any $f \in \mathcal{F}$, $i \in \{1, \dots, N\}$ with $N := \mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty)$

$$|\hat{R}_S(f) - R(f)| \leq |\hat{R}_S(f) - \hat{R}_S(f_i)| + |\hat{R}_S(f_i) - R(f_i)| + |R(f_i) - R(f)|$$

$$|\hat{R}_S(f) - \hat{R}_S(f_i)| = \left|\frac{1}{n}\sum_{j=1}^{n}[\ell(y_j, f(x_j)) - \ell(y_j, f_i(x_j))]\right| \leq L \cdot \|f - f_i\|_\infty \leq L \cdot \varepsilon$$

$$|R(f) - R(f_i)| = |\mathbb{E}[\ell(y, f(x)) - \ell(y, f_i(x))]| \leq L \cdot \mathbb{E}\,|f(x) - f_i(x)| \leq L \cdot \varepsilon.$$

Thus,

$$|\hat{R}_S(f) - R(f)| \leq 2L\cdot\varepsilon + |\hat{R}_S(f_i) - R(f_i)| \leq 2L\varepsilon + \max_{1\leq k\leq N} |\hat{R}_S(f_k) - R(f_k)|$$

Hence,

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |\hat{R}_S(f) - R(f)|\right] \leq 2L\varepsilon + \mathbb{E}\left[\max_{1\leq i\leq N} |\hat{R}_S(f_i) - R(f_i)|\right].$$

Note that $Z_i = |\hat{R}_S(f_i) - R(f_i)| \in [0, a]$ a.s. for all i.

$\Rightarrow$ From HW1, $\mathbb{E}\left[\max_{1\leq i\leq N} Z_i\right] \leq 2a\sqrt{\frac{2\log(2N)}{n}}$.

Hence,

$$\mathbb{E}\sup_{f\in\mathcal{F}} |\hat{R}_S(f) - R(f)| \leq 2L\varepsilon + 2a\sqrt{\frac{2\log(2\mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty))}{n}}.$$

■

Bounding Rademacher Complexity with Covering Number

The following theorem provides an upper bound for Rademacher complexity in terms of covering numbers.

THEOREM 2 Given $\mathcal{U} \subset \mathbb{R}^n$,

$$\mathbb{E}_\varepsilon\left[\sup_{u\in\mathcal{U}} \sum_{j=1}^{n} \varepsilon_j u_j\right] \leq \inf_{\alpha>0}\left\{\alpha\sqrt{n} + \left[\sup_{u\in\mathcal{U}} \|u\|_2\right]\sqrt{2\log \mathcal{N}(\mathcal{U}, \alpha, \|\cdot\|_2)}\right\}$$

Thus,

$$\widehat{\mathrm{Rad}}_S(\mathcal{F}) \leq \inf_{\alpha>0}\left\{\frac{\alpha}{\sqrt{n}} + \left[\sup_{f\in\mathcal{F}} \frac{\sqrt{\sum_{j=1}^{n}|f(x_j)|^2}}{\sqrt{n}}\right] \cdot \sqrt{2\log \mathcal{N}(\mathcal{F}|_S, \alpha, \|\cdot\|_2)}\right\}$$

Proof For an arbitrary $\alpha > 0$, suppose $\mathcal{N}(\mathcal{U}, \alpha, \|\cdot\|_2) < \infty$.

Let $V = \bigcap_{\substack{W\subset\mathcal{U} \\ W \text{ is an } \alpha\text{-cover}}} W$. Also, for any $u\in\mathcal{U}$, let $V(u) \in \operatorname{argmin}_{v\in V} \|u - v\|_2$

Then,

$$\mathbb{E}_\varepsilon\left[\sup_{u\in\mathcal{U}} \sum_{j=1}^{n} \varepsilon_j u_j\right] = \mathbb{E}_\varepsilon\left[\sup_{u\in\mathcal{U}} \sum_{j=1}^{n} \varepsilon_j \left(u_j - V_j(u) + V_j(u)\right)\right]$$

$$= \mathbb{E}\left[\sup_{u\in\mathcal{U}}\left\{\sum_{j=1}^{n}\varepsilon_j V_j(u) + \sum_{j=1}^{n}\varepsilon_j\left(u_j - V_j(u)\right)\right\}\right]$$

$$\overset{C.S.}{\leq} \mathbb{E}\left[\sup_{u\in\mathcal{U}}\left\{\sum_{j=1}^{n}\varepsilon_j V_j(u) + \sqrt{\sum_{j=1}^{n}\varepsilon_j^2}\sqrt{\sum_{j=1}^{n}\left(u_j - V_j(u)\right)^2}\right\}\right]$$

$$= \mathbb{E}\left[\sup_{u\in\mathcal{U}}\left\{\sum_{j=1}^{n}\varepsilon_j V_j(u) + \sqrt{n}\cdot\|u - V(u)\|_2\right\}\right]$$

$$\leq \mathbb{E}\left[\sup_{u\in\mathcal{U}}\sum_{j=1}^{n}\varepsilon_j V_j(u)\right] + \alpha\sqrt{n}$$

$$\leq \mathbb{E}\left[\sup_{v\in V}\sum_{j=1}^{n}\varepsilon_j v_j\right] + \alpha\sqrt{n}$$

$$\leq \sup_{v\in V}\|v\|_2\cdot\sqrt{2\log|V|} + \alpha\sqrt{n} \qquad \text{Massart's finite class}$$

$$\leq \sup_{u\in\mathcal{U}}\|u\|_2\cdot\sqrt{2\log|V|} + \alpha\sqrt{n} \qquad V \text{ is a proper cover}$$

The bound holds since $|V| = \mathcal{N}(\mathcal{U},\alpha,\|\cdot\|_2)$, and α is arbitrary. ■

Covering number provides a new method to bound $\widehat{\text{Rads}}$ and Radn.

Covering Number Bounds for Linear Function Classes :

THEOREM 3 (Covering number bounds for linear predictors)

Let $x_j \in \mathbb{R}^d$ with $\|x_j\|_p \leq b$ for all $j\in[n]$,

$\mathcal{F} = \{Xw : w\in\mathbb{R}^d, \|w\|_q \leq a\}$ where $X_{j,:} = x_j^T, \forall j\in\{1,\dots,n\}$,

(p,q) is s.t. $\frac{1}{p}+\frac{1}{q}=1$, $p\geq 2$. Then,

$$\log_2 \mathcal{N}(\mathcal{F},\varepsilon,\|\cdot\|_2) \leq \frac{a^2b^2}{\varepsilon^2}\cdot\log_2(2d+1).$$

<u>Proof</u> $X = \begin{bmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix} = \begin{bmatrix} | & & | \\ y_1 & \cdots & y_d \\ | & & | \end{bmatrix} \in \mathbb{R}^{n\times d}$.

Let $g_j = \dfrac{n^{\frac{1}{p}} . ab \; y_j}{\|y_j\|_p}$, $\omega_j' = \dfrac{\|y_j\|_p}{n^{\frac{1}{p}} ab} \omega_j$. Then, $\sum_j g_j \omega_j' = X\omega$.

By Hölder's inequality, check that

$$\sum_{j=1}^{d} |\omega_j'| = \sum_{j=1}^{d} \frac{1}{n^{\frac{1}{p}} ab} \cdot \|y_j\|_p \cdot |\omega_j| \leq \frac{1}{n^{\frac{1}{p}} ab} \Big(\sum_j \|y_j\|_p^p\Big)^{\frac{1}{p}} \Big(\sum_j |\omega_j|^q\Big)^{\frac{1}{q}}$$

$$\leq \frac{1}{n^{\frac{1}{p}} ab} \cdot \left(n . b^p \right)^{\frac{1}{p}} . \; a \quad = 1.$$

$z \mapsto z^{p/2}$ is convex $\Longrightarrow$ $\dfrac{1}{\sqrt{n}} \|y_j\|_2 \leq \dfrac{1}{n^{1/p}} \|y_j\|_p$ for all j

by Jensen's inequality:

$$\left(\frac{1}{n} \sum_{k=1}^{n} y_{jk}^2\right)^{\frac{p}{2}} \leq \frac{1}{n} \sum_{k=1}^{n} |y_{jk}|^p$$

$$\Longrightarrow \left(\frac{1}{n} \sum_{k=1}^{n} y_{jk}^2 \right)^{\frac{1}{2}} \leq \left(\frac{1}{n} \sum_{k=1}^{n} |y_{jk}|^p\right)^{\frac{1}{p}}.$$

Therefore, by Maurey's Lemma, if $k \geq \left(\frac{ab}{\varepsilon}\right)^2$, we can find $(k_1, \dots, k_d) \in \mathbb{Z}^d$ s.t. $\sum_{j=1}^{d} |k_j| \leq k$ and

$$\left\| z - \frac{1}{k} \sum_{j=1}^{d} k_j g_j \right\|_2^2 \leq \frac{n a^2 b^2}{k} \leq \varepsilon^2 . n$$

Thus, $\mathcal{N}(\mathcal{F}, \varepsilon, \|.\|_2) \leq \left| \left\{ (k_1, \dots, k_d) \in \mathbb{Z}^d : \sum_{j=1}^{d} |k_j| \leq k \right\} \right|$

$$\leq (2d+1)^k.$$

Lemma (Maurey) Fix a Hilbert space $\mathcal{H}$ with norm $\|.\|$.
Let $u \in \mathcal{H}$ be s.t. $u = \sum_{i=1}^{d} \alpha_i v_i$ for $\alpha_i \geq 0$, $v_i \in \mathcal{H}$ for $i \in [d]$, and $\alpha \neq 0$. Then, for any integer $k \in \mathbb{Z}_+$, $\exists (k_1, \dots, k_d) \in \mathbb{Z}_+^d$ s.t. $\sum_{i=1}^{d} k_i = k$ s.t.

$$\left\| u - \|\alpha\|_1 \cdot \frac{1}{k} \sum_{i=1}^{d} k_i v_i \right\| \leq \|\alpha\|_1 \cdot \frac{1}{k} \sum_{i=1}^{d} \alpha_i \|v_i\|^2$$

$$\leq \|\alpha\|_1^2 \cdot \max_j \|v_i\|^2 / k$$

Proof Let $\beta = \|\alpha\|_1$, let $\varphi_1, \varphi_2, \dots, \varphi_k$ denote k iid r.v.'s with

$$\mathbb{P}(\varphi_1 = \beta . v_i) = \frac{\alpha_i}{\beta}, \text{ for all } i. \text{ Define}$$

$$\overline{\varphi} = \frac{1}{k} \sum_{j=1}^{k} \varphi_j .$$

Then,

$$\mathbb{E}[\overline{\varphi}] = \frac{1}{k} \sum_{j=1}^{k} \mathbb{E}[\varphi_j] = \frac{1}{k} \sum_{j=1}^{k} \sum_{i=1}^{d} \frac{\alpha_i}{\beta} . \beta . v_i$$

$$= \sum_{i=1}^{d} \alpha_i v_i = u .$$

Consequently,

$$\mathbb{E} \|u - \overline{\varphi}\|^2 = \frac{1}{k^2} \mathbb{E} \left\| \sum_{j=1}^{k} (\varphi_j - u) \right\|^2$$

$$= \frac{1}{k^2} \mathbb{E} \left[\sum_{j=1}^{k} \|\varphi_j - u\|^2 + \sum_{i \neq j} \langle u - \varphi_i, u - \varphi_j \rangle \right]$$

$$= \frac{1}{k} \mathbb{E}[\|\varphi_1 - u\|^2] = \frac{1}{k} \mathbb{E}[\|\varphi_1 - \mathbb{E}[\varphi_1]\|^2]$$

$$\leq \frac{1}{k} \mathbb{E} \|\varphi_1\|^2 = \frac{1}{k} \sum_{i=1}^{d} \frac{\alpha_i}{\beta} . \beta^2 . \|v_i\|^2 = \frac{\beta}{k} \sum_{i=1}^{d} \alpha_i \|v_i\|^2 .$$

To conclude, $\exists (j_1, \dots, j_k) \in \{1, \dots, d\}^k$ and an assignment $\hat{\varphi}_i := \beta v_{j_i}$ and $\hat{\varphi} = \frac{1}{k} \sum_{i=1}^{k} \hat{\varphi}_i$ s.t. $\|u - \hat{\varphi}\|^2 \leq \mathbb{E} \|u - \overline{\varphi}\|^2$.