1 Recap

1. Covering: $V$ is an $\ell_p$-cover of $\mathcal{F}$ on $x_1, \ldots, x_n$ at scale $\beta$ if

$$\forall f \in \mathcal{F}, \exists v \in V \text{ s.t. } \left(\frac{1}{n} \sum_{t=1}^{n} |f(x_t) - v[t]|^p\right)^{1/p} \leq \beta$$

$$\mathcal{N}_p(\mathcal{F}, \beta; x_1, \ldots, x_n) = \min\{|V| : V \text{ is an } \ell_p\text{-cover of } \mathcal{F} \text{ on } x_1, \ldots, x_n \text{ at scale } \beta\}$$

2. $E_S\left[L_D(\hat{y}_{\text{erm}}) - \inf_{f \in \mathcal{F}} L_D(f)\right] \leq 2E_S\left[\hat{R}_S(\mathcal{F})\right] \leq 2 \inf_{\beta > 0} \left\{\beta + \sqrt{\frac{\log N_1(\mathcal{F}, \beta; x_1, \ldots, x_n)}{n}}\right\}$

3. $\hat{R}_S(\mathcal{F}) \leq \hat{D}_S(\mathcal{F}) := \inf_{\alpha > 0} \left\{4\alpha + 12 \int_{\alpha}^{1} \sqrt{\frac{\log N_2(\mathcal{F}, \beta; x_1, \ldots, x_n)}{n}} d\beta\right\}$

Also, $\hat{R}_S(\mathcal{F}) \geq \tilde{\Omega}\left(\hat{D}_S(\mathcal{F})\right)$

2 Fat Shattering Dimension

Definition 1. We say that $\mathcal{F}$ shatters $x_1, \ldots, x_n$ at scale $\gamma$, if there exists witness $s_1, \ldots, s_n$ such that, for every $\epsilon \in \{-1\}^n$, there exists $f_\epsilon \in \mathcal{F}$ such that

$$\forall t \in [n], \quad \epsilon_t \cdot (f_\epsilon(x_t) - s_t) \geq \gamma/2$$

Further $\text{fat}_\gamma(\mathcal{F}) = \max\{n : \exists x_1, \ldots, x_n \in \mathcal{X} \text{ s.t. } \mathcal{F} \text{ $\gamma$-shatters } x_1, \ldots, x_n\}$

\[\text{Diagram:\ }] \text{A graph showing a function and its witness points.}\]
Theorem 1. For any $\mathcal{F} \subseteq [-1,1]^X$ and any $\gamma \in (0,1)$

$$N_2(\mathcal{F}, \gamma, n) \leq \left( \frac{2}{\gamma} \right)^{K \ fat_{e\gamma}(\mathcal{F})}$$

where in the above $c$ and $K$ are universal constants.

Using the above with the dudley chaining bounds we get,

$$D_S(\mathcal{F}) \leq \inf_{\alpha \geq 0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_0^1 \sqrt{K \ fat_{e\delta}(\mathcal{F}) \log \left( \frac{2}{\delta} \right)} d\delta \right\}$$

Thus bound on fat-shattering dimension leads to bound on Rademacher complexity.

Binary function class

For any $\delta \in [0,1)$, and any $c \leq 1$, $fat_{c\delta}(\mathcal{F}) = fat_0(\mathcal{F}) = VC(\mathcal{F})$ we can conclude that $\gamma^{stat}_n(\mathcal{F}) \leq R_n(\mathcal{F}) \leq \sqrt{\frac{VC(\mathcal{F})}{n}}$.

Linear Predictors

Let $X = \{ x : \|x\|_2 \leq 1 \}$ and let $\mathcal{F} = \{ x \mapsto f^\top x : \|f\|_2 \leq 1 \}$.

1. $fat_{\gamma}(\mathcal{F}) \geq \lfloor 4\gamma^{-2} \rfloor$:

For all $i \in [d]$, let $x_i = e_i$ and let $s_i = 0$. Given $\epsilon \in \{\pm 1\}^d$, consider the vector $f$ such that $f[i] = \epsilon_i \gamma/2$. Clearly $f$, $\gamma$-shatters these set of $d$ points. Now for $\|f\|_2 \leq 1$, we need that $\sum_{i=1}^d f^2[i] = d\gamma^2/4 \leq 1$. This implies that $d \leq 4\gamma^{-2}$. Thus we can provide $4/\gamma^2$ points that can be $\gamma$-shattered.

2. $fat_{\gamma}(\mathcal{F}) \leq 4\gamma^{-2}$:

Typically uses Maurey’s theorem but we will take a different route in just a bit.

2.1 Back to Rademacher

Claim 2.

$$R_n(\mathcal{F}) \geq \sup \{ \gamma/2 : fat_{\gamma}(\mathcal{F}) > n \}$$

Proof. Think about Rademacher complexity on shattered points. □

The claim above is the same as saying (converse) $fat_{\gamma} \leq \min \{ n : R_n(\mathcal{F}) \leq \gamma/2 \}$. Using this for linear class example, since we know that $R_n(\mathcal{F}) \leq \frac{1}{\sqrt{n}}$, we can conclude that for the linear class, $fat_{\gamma} \leq \min \{ n : R_n(\mathcal{F}) \leq \gamma/2 \} \leq \min \{ n : \frac{1}{\sqrt{n}} \leq \gamma/2 \} \leq \lceil \frac{4}{\gamma^2} \rceil$.

Using a more refined argument, the claim above can be improved, it can be shown that for any $\gamma > R_n(\mathcal{F})$,

$$fat_{\gamma}(\mathcal{F}) \leq \frac{8nR_n^2(\mathcal{F})}{\gamma^2}$$

from this we can conclude that

$$\hat{R}_S(\mathcal{F}) \geq \Omega \left( \inf_{\alpha \geq 0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_0^1 \sqrt{K \ fat_{e\delta}(\mathcal{F}) \log \left( \frac{2}{\delta} \right)} d\delta \right\} \right)$$