3.9&3.10 Massart's finite Lemma and shattering coefficient

[Lemma 5] Massart's finite lemma.

Throughout this lemma, condition on $Z_1, \cdots, Z_n$.

Let F be a finite set of functions. Let M^2 be the const. s.t.

$$\sup_{f\in F} \frac{1}{n}\sum_{i=1}^n f(Z_i)^2 \le M^2$$

Then the empirical Rademacher complexity is upper bounded by: $\hat{R}_n(F) \le \sqrt{2M^2 \log|F| / n}$

Pf: Let $W_f = \frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)$, then

$$\hat{R}_n(F) = E[\sup_{f\in F} W_f \mid Z_{1:n}]$$

$$\Rightarrow \exp\{t\hat{R}_n(F)\} \le E[\exp\{t\sup_{f\in F} W_f\} \mid Z_{1:n}] \quad \text{Jensen's ineq.}$$

$$= E[\sup_{f\in F} \exp\{tW_f\} \mid Z_{1:n}] \quad \text{monotonicity}$$

$$\le \sum_{f\in F} E[\exp\{tW_f\} \mid Z_{1:n}]$$

By Hoeffding's Lemma,

σ_i is sub-Gaussian with parameter 1,

W_f is sub-Gaussian with parameter $\frac{1}{n^2}\sum_{i=1}^n f(Z_i)^2 \le \frac{M^2}{n}$

since Z_i are fixed and $\{\sigma_i\}$ are independent.

Then: $E[\exp\{tW_f\}] \le \exp\{\frac{t^2M^2}{2n}\}$

$$\Rightarrow \exp\{t\hat{R}_n(F)\} \le \sum_{f\in F} E[\exp\{tW_f\} \mid Z_{1:n}]$$

$$\le |F| \exp\{\frac{t^2M^2}{2n}\}$$

$$\Rightarrow \hat{R}_n(F) \le \frac{\log|F|}{t} + \frac{tM^2}{2n} \quad \text{for any } t>0$$

$$\Rightarrow \hat{R}_n(F) \le 2\sqrt{\log|F|M^2/2n} \qquad \square$$

Questions: How $\hat{R}_n(F)$ and $R_n(F)$ differ?

- To simplify the estimate, let's first reward them as the same. Maybe we could answer it in the perspective of ghost dataset.

3.10. Shattering coefficient

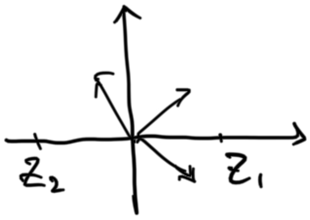
Define $\mathcal{F} = \{ z \mapsto \mathbb{1}\{w \cdot z \geq 0\} : w \in \mathbb{R}^d \}$, then $\mathcal{F}$ is infinite.

Question: What should the complexity of $\mathcal{F}$ be?

Idea: find equivalent finite $\mathcal{F}'$ w.r.t. $Z_{1:n}$.

example: consider $Z_1 = (3, 0)$, $Z_2 = (-3, 0)$, define

$\mathcal{F}' := \{ z \mapsto \mathbb{1}\{z \geq 0\}, z \mapsto \mathbb{1}\{-z \geq 0\} \}$ has the same behaviors as $\mathcal{F}$.



[Definition 10] shattering coefficient

Let $\mathcal{F}$ be a family of functions that map $\mathcal{Z}$ to a finite set (usually $\{0,1\}$). The shattering coefficient of $\mathcal{F}$ is the maximum number of behaviors over n points:

$$s(\mathcal{F}, n) := \max_{z_1, \cdots, z_n \in \mathcal{Z}} \left| \{ [f(z_1), \cdots, f(z_n)] : f \in \mathcal{F} \} \right|$$

Since the shattering coefficient is finite (by def), by Lemma 5, we have

$$\hat{R}_n(\mathcal{F}) \leq \sqrt{2 \log s(\mathcal{F}, n) / n}$$

Note: (i) for one-zero classes, $s(\mathcal{F}, n) \leq 2^n$. if $s(\mathcal{F}, n) = 2^n$, $\mathcal{F}$ is said to shatter any n points.

(ii) $s(\mathcal{F}, n) \sim$ sub-exponential is meaningful

(iii) hypothesis class $\mathcal{H} = \{ \mathcal{X} \to \{0,1\} \}$,

loss class: $\mathcal{A} = \{ (x,y) \to \mathbb{1}\{y \neq h(x)\} : h \in \mathcal{H} \}$

we have $s(\mathcal{H}, n) = s(\mathcal{A}, n)$.

(iii) is because

$$[\ell((x_i, y_i), h)]_{i=1}^{n} \iff [h(x_1), \cdots, h(x_n)]$$