- Rademacher Random Variable

  ε_i is a Rademacher r.v.

  $\mathbb{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$

  $\mathbb{P}(\varepsilon_i = k) = 0 \quad \forall k \neq \pm 1$

- Rademacher Complexity

  $R_n(\mathcal{F}) := \mathbb{E}[\sup_{f \in \mathcal{F}} n^{-1} \sum_{i=1}^n \varepsilon_i f(x_i)]$

We first look at R.C. of linear function classes

$(x \sim N(0, I))$

$\mathcal{F}_p := \{x \mapsto \langle x, w \rangle \mid \|w\|_p \le 1\}$

Can use dual norm $\|\cdot\|_*$

$\|\cdot\| = \|\cdot\|_p$

$\|\cdot\|_* = \|\cdot\|_q$

$\|x\|_* := \sup_{\|z\| \le 1} \langle z, x \rangle$

$\frac{1}{p} + \frac{1}{q} = 1$

$R_n(\mathcal{F}_p) = \mathbb{E}[\sup_{\|w\|_p \le 1} n^{-1} \sum_{i=1}^n \varepsilon_i \langle x_i, w \rangle$

$= n^{-1} \mathbb{E} \sup_{\|w\|_p \le 1} \langle w, \sum_{i=1}^n \varepsilon_i x_i \rangle$

$= n^{-1} \mathbb{E} \|\sum_{i=1}^n \varepsilon_i x_i\|_q$

where $\frac{1}{p} + \frac{1}{q} = 1$

$\frac{1}{\sqrt{n}} \sum_{j=1}^d \mathbb{E} |\frac{Z_j}{\sqrt{n}}| = \sqrt{\frac{2}{\pi}}$

$\frac{1}{\sqrt{n}} \mathbb{E}[\sum_{j=1}^d |\frac{Z_j}{\sqrt{n}}|]$

(i) $p = \infty$

$R_n(\mathcal{F}_\infty) = n^{-1} \mathbb{E} \|\sum_{i=1}^n \varepsilon_i x_i\|_1 = \mathbb{E} \left\| \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} \right\|_1$

$\underline{Z_j} := \sum_{i=1}^n \varepsilon_i \langle x_i, e_j \rangle \overset{d}{=} N(0, n)$

$\Rightarrow R_n(\mathcal{F}_\infty) = \frac{\boxed{d}}{n} \sqrt{n} \sqrt{\frac{2}{\pi}} = \sqrt{\frac{2}{\pi}} \cdot \frac{d}{\sqrt{n}} \Leftarrow \frac{d}{\sqrt{n}} \cdot \sqrt{\frac{2}{\pi}}$

$$E\left|\underbrace{\frac{Z_j}{\sqrt{n}}}_{\sim N(0,1)}\right| = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |s| \exp\left(-\frac{s^2}{2}\right) ds$$

$$= \frac{2}{\sqrt{2\pi}}\int_0^\infty s \exp\left(-\frac{s^2}{2}\right) ds$$

$$\left(u = \frac{s^2}{2} \Rightarrow du = s\,ds\right)$$

$$= \sqrt{\frac{2}{\pi}}\int_0^\infty \exp(-u)\, du = \sqrt{\frac{2}{\pi}}$$

(2) $p=1$

$$R_n(\mathcal{F}_1) = n^{-1} E\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_\infty$$

Exercise B.18

Let X_i, $i=1,\cdots,n$ be zero-mean σ-sub-Gaussian r.v.. We have

$$E \max_{i=1,\cdots,n} |X_i| \leq \sigma\sqrt{2\log(2n)}$$

$$E\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_\infty = E \max_{j=1,\cdots,d} |Z_j| \leq \sqrt{n}\sqrt{2\log(2d)}$$

$$\Rightarrow R_n(\mathcal{F}_1) \leq \sqrt{\frac{2\log(2d)}{n}}$$

(iii) $p=2$

$$\mathcal{R}_n(\mathcal{F}_2) \le \sqrt{\frac{d}{n}}$$

Exercise 1.36:

WTS: $\mathcal{R}_n(\mathcal{F}_2) \geq c\sqrt{\frac{d}{n}}$ for a univ. const. c

Proposition 1.41:

Let $\phi(x) = \max\{x, 0\}$ be the RELU act. func.

$$\mathcal{F} = \{x \mapsto \textstyle\sum_{h=1}^{m} w_h \phi(\langle u_h, x\rangle) \mid \{w_h\}_{h=1}^m \subset \mathbb{R}, \{u_h\}_{h=1}^m \subset \mathbb{R}^d, \underbrace{\textstyle\sum_{h=1}^m |w_h| \|u_h\| \le 1}\}$$

$$\Rightarrow \mathcal{R}_n(\mathcal{F}) \le 2\sqrt{\frac{d}{n}}$$

Lemma: Let $\ell: \mathbb{R} \mapsto \mathbb{R}$ be L-Lipschitz:

$$\forall x, y \in \mathbb{R},\ |\ell(x) - \ell(y)| \le L|x-y|$$

Let $\mathcal{F}$ be a function class of $\underbrace{X \mapsto \mathbb{R}}$

$$\mathcal{R}_n(\{x \mapsto \ell(f(x)) \mid f \in \mathcal{F}\}) \le L \cdot \mathcal{R}_n(\mathcal{F})$$

$\rightarrow$ Proposition 1.29

$$\theta = \{\{w_h\}, \{u_h\}\}$$

Proof:
$$n \cdot \mathcal{R}_n(\mathcal{F}) = \mathbb{E} \sup_{\theta \in \Theta} \sum_{i=1}^{n} \varepsilon_i \sum_{h=1}^{m} w_h \phi(\langle u_h, x_i\rangle)$$
$$= \mathbb{E} \sup_{\theta \in \Theta} \sum_{h=1}^{m} w_h \textstyle\sum_{i=1}^{n} \varepsilon_i \phi(\langle u_h, x_i\rangle)$$
$$= \mathbb{E} \sup_{\theta \in \Theta} \underbrace{\textstyle\sum_{h=1}^{m} w_h \|u_h\|} \textstyle\sum_{i=1}^{n} \varepsilon_i \phi(\langle u_h, x_i\rangle / \|u_h\|)$$

$$\le \mathbb{E} \sup_{\theta \in \Theta} \sum_{h=1}^m w_h \|u_h\| \cdot \max_{h=1,\cdots,m} \left| \sum_{i=1}^n \varepsilon_i \phi(\langle u_h, x_i \rangle / \|u_h\|) \right|$$

$$\le \mathbb{E} \sup_{\theta \in \Theta} \max_{h=1,\cdots,m} \left| \sum_{i=1}^n \varepsilon_i \phi(\langle u_h, x_i \rangle / \|u_h\|) \right|$$

$$\le \mathbb{E} \sup_{\|u\| \le 1} \cancel{\max_{h=1,\cdots,m}} \left| \sum_{i=1}^n \varepsilon_i \phi(\langle u, x_i \rangle) \right|$$

$$= \mathbb{E} \sup_{\|u\| \le 1} \left| \sum_{i=1}^n \varepsilon_i \phi(\langle u, x_i \rangle) \right|$$

Prop. 1.42. Let $T \subset \mathbb{R}^n$ be an arbitrary set and suppose that:

$$\forall \varepsilon \in \{\pm 1\}^n, \ \exists t \in T \ \text{ s.t. } \langle \varepsilon, t \rangle \ge 0$$

Then $\mathbb{E} \sup_{t \in T} |\langle \varepsilon, t \rangle| \le 2 \mathbb{E} \sup_{t \in T} \langle \varepsilon, t \rangle$ where $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)$ and the ε_i's are i.i.d. Rademacher r.v.

$Q(x_{1:n}) \subset \mathbb{R}^n$ s.t.

$$Q(x_{1:n}) := \{ z = (\phi(\langle u, x_i \rangle))_{i=1}^n \mid \|u\| \le 1 \}$$

$0 \in Q(x_{1:n})$

$$n \cdot \mathcal{R}_n(\mathcal{F}) \le \mathbb{E} \sup_{\|u\| \le 1} \left| \sum_{i=1}^n \varepsilon_i \phi(\langle u, x_i \rangle) \right|$$
$$= \mathbb{E}_{x_{1:n}} \mathbb{E}_\varepsilon \sup_{q \in Q(x_{1:n})} |\langle \varepsilon, q \rangle|$$

$$\leq 2 \mathbb{E}_{x_{1:n}} \mathbb{E}_{\varepsilon} \sup_{q \in Q(x_{1:n})} \langle \varepsilon, q \rangle$$

$$= 2 \mathbb{E} \sup_{\|u\| \leq 1} \sum_{i=1}^{n} \varepsilon_i \phi(\langle u, x_i \rangle)$$

$$\leq 2 \mathbb{E} \sup_{\|u\| \leq 1} \sum_{i=1}^{n} \varepsilon_i \langle u, x_i \rangle$$

$$= 2\sqrt{nd}$$

$$\Rightarrow \mathcal{R}_n(\mathcal{F}) \leq 2\sqrt{\frac{d}{n}}$$