1 Rademacher complexity of ℓ_1 linear class

$\mathcal{H}_1 = \{w^T x : \|w\|_1 \le 1\}$ Suppose $\|x_i\|_\infty \le 1$, $i=1,\cdots n$

(a) Show that $\widehat{\text{Rad}}_n(\mathcal{H}_1) = \frac{1}{n} \mathbb{E}_\xi \| \sum_{i=1}^n \xi_i X_i \|_\infty$

Proof.
$$\widehat{\text{Rad}}_n(\mathcal{H}_1) = \mathbb{E}_\xi \sup_{\|w\|_1 \le 1} \frac{1}{n} \sum_{i=1}^n \xi_i w^T X_i = \mathbb{E}_\xi \sup_{\|w\|_1 \le 1} w^T \left(\frac{1}{n} \sum_{i=1}^n \xi_i X_i\right)$$

$$\le \mathbb{E}_\xi \| \frac{1}{n} \sum_{i=1}^n \xi_i X_i \|_\infty = \frac{1}{n} \mathbb{E}_\xi \| \sum_{i=1}^n \xi_i X_i \|_\infty$$

as $|a^T b| \le \|a\|_1 \|b\|_\infty$. #

(b) Show that $\widehat{\text{Rad}}_n(\mathcal{H}_1) \le \sqrt{\frac{2\log(2d)}{n}}$

Proof. $Z = \frac{1}{n} \sum_{i=1}^n \xi_i X_i$ is d-dimensional random vector

$Z_j = \frac{1}{n} \sum_{i=1}^n \xi_i X_{ij}$, $j=1,\cdots d$ $\quad \mathbb{E} Z_j = 0$

$$\psi(\lambda) = \log \mathbb{E} e^{\lambda(Z_j - \mathbb{E} Z_j)} = \log \mathbb{E} e^{\lambda Z_j}$$

$$= \log \mathbb{E} e^{\frac{\lambda}{n} \sum_{i=1}^n X_{ij} \xi_i} = \log \prod_{i=1}^n \mathbb{E} e^{\frac{\lambda}{n} X_{ij} \xi_i}$$

$$= \sum_{i=1}^n \log \mathbb{E} e^{\frac{\lambda}{n} X_{ij} \xi_i} = \sum_{i=1}^n \log \left(\frac{1}{2} e^{\frac{\lambda}{n} X_{ij}} + \frac{1}{2} e^{-\frac{\lambda}{n} X_{ij}}\right)$$

$$\le \sum_{i=1}^n \log\left(\frac{e^{\frac{\lambda}{n}} + e^{-\frac{\lambda}{n}}}{2}\right) \le \sum_{i=1}^n \log \frac{e^\lambda + e^{-\lambda}}{2} \le \sum_{i=1}^n \log e^{\frac{\lambda^2}{2}} = \frac{\lambda^2 n}{2}$$

This shows Z_j is sub-gaussian variable with proxy $\sigma^2 = n$. Same holds for $-Z_j$.

By maximal inequality on $\{Z_j\}_{j=1}^d$, $\{-Z_j\}_{j=1}^d$

$$\mathbb{E} \max_{j \in [d]} \{Z_j, -Z_j\} \le \sqrt{2n \log(2d)} .$$

Thus $\widehat{\text{Rad}}_n(\mathcal{H}_1) = \frac{1}{n} \mathbb{E}_\xi \|Z\|_\infty = \frac{1}{n} \mathbb{E} \max_{j \in [d]} \{Z_j, -Z_j\}$

$$\le \sqrt{\frac{2\log(2d)}{n}} \quad \#$$

2 Empirical method of Maurey

$S = \{x \in \mathbb{R}^d \mid \sum_{j=1}^d x_j = 1,\ x_j \geq 0\}$

For $x \in S$, define $P_x(z = e_j) = x_j$, $j = 1, \dots d$

(a) $z_1 \cdots z_m \underset{\sim}{\text{iid}} \mathbb{P}_x$ $\quad \bar{z} = \frac{1}{m} \sum_{j=1}^m z_j$ $\quad$ Prove $\mathbb{E}_{z_1 \cdots z_m} \| \bar{z} - x \|_2^2 \leq \frac{1}{m}$

Proof. Let $x = (x_1, \cdots x_d)^T$ We have $\mathbb{E}_{z_1 \cdots z_m} \bar{z} = \mathbb{E}_x z = \sum_{j=1}^m x_j e_j = x$

$$
\begin{aligned}
\mathbb{E}_{z_1 \cdots z_m} \| \bar{z} - x \|_2^2 &= \mathbb{E}_{z_1 \cdots z_m} (\bar{z}^T - x^T)(\bar{z} - x) \\
&= \mathbb{E}_{z_1 \cdots z_m} \bar{z}^T \bar{z} - 2x^T \mathbb{E}_{z_1 \cdots z_m} \bar{z} + x^T x \\
&= \mathbb{E}_{z_1 \cdots z_m} \bar{z}^T \bar{z} - x^T x \\
&= \frac{1}{m^2} \mathbb{E}_{z_1 \cdots z_m} \left(\sum_{i=1}^m z_i^T z_i + 2 \sum_{i<j} z_i^T z_j \right) - \sum_{j=1}^d x_j^2 \\
&= \frac{1}{m} \mathbb{E}_{z \sim x} z^T z + \frac{m-1}{m} \| \mathbb{E}_{z \sim x} z \|^2 - \sum_{j=1}^d x_j^2 \\
&= \frac{1}{m} + \frac{m-1}{m} \sum_{j=1}^d x_j^2 - \sum_{j=1}^d x_j^2 \\
&= \frac{1}{m} - \frac{1}{m} \sum_{j=1}^d x_j^2 \leq \frac{1}{m} \quad .\#
\end{aligned}
$$

(b) $m \in \mathbb{N}_+$ $\quad \mathcal{N}_m \subseteq S$ be the set of all possible $\bar{z}$

Show that $\mathcal{N}_m = \{ \frac{1}{m} \sum_{j=1}^d a_j e_j : \sum_{j=1}^d a_j = m,\ a_j \in \mathbb{N}_0,\ \forall j \in [d] \}$

and $|\mathcal{N}_m| \leq d^m$

Proof. From definition of $\mathbb{P}_x$,

$$
\begin{aligned}
\mathcal{N}_m &= \{ \frac{1}{m} \sum_{j=1}^m z_j \mid z_j \in \{e_1, \cdots e_d\},\ j = 1, \cdots m \} \\
&= \{ \frac{1}{m} \sum_{j=1}^d a_j e_j \mid \sum_{j=1}^d a_j = m,\ a_j \in \mathbb{N},\ a_j \geq 0,\ \forall j \in [d] \}
\end{aligned}
$$

Because $\# \{ a_j : a_j \in \mathbb{N},\ a_j \geq 0,\ \sum_{j=1}^d a_j = m \} = \binom{m+d-1}{d-1}$

$$
= \frac{(m+d-1)!}{(d-1)!\, m!} = \frac{(m+d-1) \cdots (d+1) \cdot d}{m \cdot (m-1) \cdots 2 \cdot 1} \leq d^m
$$

It follows $|\mathcal{N}_m| \leq d^m$ #

(c) For any $\varepsilon \in (0,1)$, $m_\varepsilon = \lceil \frac{1}{\varepsilon} \rceil$ $\mathcal{N}_{m_\varepsilon}$ is ε-cover of S

and $\log \mathcal{N}(S, \|\cdot\|_2, \varepsilon) \lesssim \frac{1}{\varepsilon^2} \log d$.

<u>Proof</u>. Because $\mathbb{E}_{z_1 \cdots z_{m_{\varepsilon^2}}} \| \bar{z} - x \|_2^2 \le \frac{1}{m_{\varepsilon^2}} = \frac{1}{\lceil \frac{1}{\varepsilon^2} \rceil} \le \varepsilon^2$

and $\mathcal{N}_{m_{\varepsilon^2}}$ is a finite set,

there exists $z_x \in \mathcal{N}_{m_{\varepsilon^2}}$, s.t. $d(z_x, x) = \| z_x - x \|_2 \le \frac{1}{\sqrt{m_{\varepsilon^2}}} \le \varepsilon$

for any $x \in S$. Thus $\mathcal{N}_{m_{\varepsilon^2}}$ is a ε-cover

$|\mathcal{N}_{m_{\varepsilon^2}}| \le d^{m_{\varepsilon^2}} = d^{\lceil \frac{1}{\varepsilon^2} \rceil} \le d^{\frac{1}{\varepsilon^2}+1}$

By definition, $\log \mathcal{N}(S, \|\cdot\|_2, \varepsilon) \le \log |\mathcal{N}_{m_{\varepsilon^2}}| = (\frac{1}{\varepsilon^2}+1) \log d$

$\le \frac{2}{\varepsilon^2} \log d$. #

3. Problem 3

$\mathcal{F}, \mathcal{G}$ function class on $\mathcal{X}$ $\quad \sup_{f,x} |f| \le A, \quad \sup_{g,x} |g| \le B$

$\mathcal{F} * \mathcal{G} = \{ fg : \mathcal{X} \to \mathbb{R} \mid f \in \mathcal{F}, g \in \mathcal{G} \}$

Show $\widehat{Rad}_n (\mathcal{F} * \mathcal{G}) \le (A+B)(\widehat{Rad}_n(\mathcal{F}) + \widehat{Rad}_n(\mathcal{G}))$

Proof. $\widehat{Rad}_n(\mathcal{F} * \mathcal{G}) = \frac{1}{n} \mathbb{E}_\xi \sup_{(f,g) \in \mathcal{F} * \mathcal{G}} \sum_{i=1}^n \xi_i f(x_i) g(x_i)$

$$= \frac{1}{4n} \mathbb{E}_\xi \sup_{\substack{f \in \mathcal{F} \\ g \in \mathcal{G}}} \sum_{i=1}^n \xi_i \left[(f(x_i) + g(x_i))^2 - (f(x_i) - g(x_i))^2 \right]$$

Note that $\pm t^2$ is $2C$-Lipschitz-continuous for $|t| \le C$.

$$\widehat{Rad}_n(\mathcal{F} * \mathcal{G}) = \frac{1}{4n} \mathbb{E}_\xi \sup_{f_1 \in \mathcal{F} + \mathcal{G}} \sum_{i=1}^n \xi_i f_1^2(x_i) + \frac{1}{4n} \mathbb{E}_\xi \sup_{f_2 \in \mathcal{F} - \mathcal{G}} \sum_{i=1}^n \xi_i (-f_2^2(x_i))$$

$$\le \frac{A+B}{2} \left[\widehat{Rad}_n(\mathcal{F} + \mathcal{G}) + \widehat{Rad}_n(\mathcal{F} - \mathcal{G}) \right] \quad (*)$$

$$\le \frac{A+B}{2} \left(\widehat{Rad}_n(\mathcal{F}) + \widehat{Rad}_n(\mathcal{G}) + \widehat{Rad}_n(\mathcal{F}) + \widehat{Rad}_n(-\mathcal{G}) \right) \quad (\dagger)$$

$$= (A+B)(\widehat{Rad}_n(\mathcal{F}) + \widehat{Rad}_n(\mathcal{G})) \quad (\triangle)$$

where $\mathcal{F} + \mathcal{G} = \{ f+g \mid f \in \mathcal{F}, g \in \mathcal{G} \}$, $\mathcal{F} - \mathcal{G} = \{ f-g \mid f \in \mathcal{F}, g \in \mathcal{G} \}$

$(*)$ follows from Contraction inequality where

$$|f(x_i) \pm g(x_i)| \le A+B$$

$(\dagger)$ proof: $\widehat{Rad}_n(\mathcal{F}_1 + \mathcal{F}_2) = \frac{1}{n} \mathbb{E}_\xi \sup_{\substack{f_1 \in \mathcal{F}_1 \\ f_2 \in \mathcal{F}_2}} \sum_{i=1}^n \xi_i [f_1(x_i) + f_2(x_i)]$

$$\le \frac{1}{n} \mathbb{E}_\xi \sup_{f_1 \in \mathcal{F}_1} \sum_{i=1}^n \xi_i f_1(x_i) + \sup_{f_2 \in \mathcal{F}_2} \sum_{i=1}^n \xi_i f_2(x_i)$$

$$= \widehat{Rad}_n(\mathcal{F}_1) + \widehat{Rad}_n(\mathcal{F}_2)$$

$(\triangle)$ follows from $\widehat{Rad}_n(-\mathcal{G}) = \widehat{Rad}_n(\mathcal{G})$

We are done. $\#$