On Bounding Radamacher Complexities

Behrad Moniri
University of Pennsylvania
bemoniri@seas.upenn.edu

Abstract

In this note, we prove various properties of Radamacher complexities. We also present bounds for Radamacher complexity of various hypothesis classes.

1 Radamacher Complexity

The empirical Radamacher complexity of a class of real-valued functions $\mathcal{F}$, given the set of instances $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is defined as

$$\mathcal{R}_S(\mathcal{F}) := \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\Big[\sup_{f\in\mathcal{F}}\sum_{i=1}^n \sigma_i f(\mathbf{x}_i)\Big].$$

We define the (population) Radamacher complexity of class $\mathcal{F}$ as

$$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_n}\Big[\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\Big[\sup_{f\in\mathcal{F}}\sum_{i=1}^n \sigma_i f(\mathbf{x}_i)\Big]\Big].$$

We also define a absolute valued Radamacher complexity of class $\mathcal{F}$ as

$$\widetilde{\mathcal{R}}_n(\mathcal{F}) = \mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_n}\Big[\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\Big[\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \sigma_i f(\mathbf{x}_i)\Big|\Big]\Big].$$

2 Basic Properties of Radamacher Complexity

In this section, we will prove several basic properties of Radamacher complexity. Let $\mathcal{F}$ and $\mathcal{G}$ be arbitrary hypothesis classes.

Theorem 1. *Let $\mathcal{F} \subseteq \mathcal{G}$, then $\mathcal{R}_n(\mathcal{F}) \leq \mathcal{R}_n(\mathcal{G})$.*

Proof. For any fixed $\boldsymbol{\sigma}$, we have

$$\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n \sigma_i f(\mathbf{x}_i) \leq \sup_{g\in\mathcal{G}}\frac{1}{n}\sum_{i=1}^n \sigma_i g(\mathbf{x}_i),$$

because $\mathcal{F} \subseteq \mathcal{G}$. Hence

$$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n \sigma_i f(\mathbf{x}_i)\Big] \leq \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{g\in\mathcal{G}}\frac{1}{n}\sum_{i=1}^n \sigma_i g(\mathbf{x}_i)\Big] = \mathcal{R}_n(\mathcal{G}).$$

which concludes the proof. □

Theorem 2. *For any $\alpha \in \mathbb{R}$, $\mathcal{R}_n(\alpha\mathcal{F}) = |\alpha|\mathcal{R}_n(\mathcal{F})$.*

Proof. The proof follows easily from the definition

$$\mathcal{R}_n(\alpha\mathcal{F}) = \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n \sigma_i \alpha f(\mathbf{x}_i)\Big] = |\alpha|\,\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n \sigma_i f(\mathbf{x}_i)\Big] = |\alpha|\,\mathcal{R}_n(\mathcal{F}).$$

the second equality follows from the fact that $-\sigma_i$ and σ_i have the same distribution. □

Theorem 3. $\mathcal{R}_n(\mathcal{F}) = \mathcal{R}_n\Big(\text{conv}(\mathcal{F})\Big)$.

Proof. Based on theorem (1), we have $\mathcal{R}_n(\mathcal{F}) \le \mathcal{R}_n\big(\text{conv}(\mathcal{F})\big)$. Based on the definition of Radamacher complexity,

$$\begin{aligned}\mathcal{R}_n\Big(\text{conv}(\mathcal{F})\Big) &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\text{conv}(\mathcal{F})} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(\mathbf{x}_i)\Big] \\ &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f_j\in\mathcal{F},\ \sum\alpha_i=1} \frac{1}{n}\sum_{i=1}^{n}\sigma_i \sum_j \alpha_j f_j(\mathbf{x}_i)\Big].\end{aligned}$$

The maximum of a linear program occurs at a corner point. Hence

$$\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f_j\in\mathcal{F},\ \sum\alpha_i=1} \frac{1}{n}\sum_{i=1}^{n}\sigma_i \sum_j \alpha_j f_j(\mathbf{x}_i)\Big] = \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f_k\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f_k(\mathbf{x}_i)\Big] = \mathcal{R}(\mathcal{F}),$$

which concludes the proof. □

Theorem 4. $\mathcal{R}_n(\mathcal{F}+\mathcal{G}) \le \mathcal{R}_n(\mathcal{F}) + \mathcal{R}_n(\mathcal{G})$.

Proof. Using the basic property $\sup(a+b) \le \sup(a) + \sup(b)$, we can write

$$\begin{aligned}\mathcal{R}_n(\mathcal{F}+\mathcal{G}) &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{h\in\mathcal{F}+\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(\mathbf{x}_i)\Big] \\ &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F},\ g\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i\big(f(\mathbf{x}_i)+g(\mathbf{x}_i)\big)\Big] \\ &\le \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(\mathbf{x}_i))\Big] + \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i g(\mathbf{x}_i))\Big] \\ &= \mathcal{R}_n(\mathcal{F}) + \mathcal{R}_n(\mathcal{G}).\end{aligned}$$

This property is indeed tight! □

Theorem 5. $\widetilde{\mathcal{R}}_n(\mathcal{F}+\{g\}) \le \tilde{\mathcal{R}}_n(\mathcal{F}) + \frac{||g||_\infty}{\sqrt{n}}$.

Proof. Based on the definition of Radamacher complexity:

$$\begin{aligned}\widetilde{\mathcal{R}}_n\big(\mathcal{F}+\{g\}\big) &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{h\in\mathcal{F}+\{g\}} \Big|\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(\mathbf{x}_i)\Big|\Big] \\ &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F}} \Big|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\big(f(\mathbf{x}_i)+g(\mathbf{x}_i)\big)\Big|\Big] \\ &\le \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{f\in\mathcal{F}} \Big|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(\mathbf{x}_i)\Big|\Big] + \mathbb{E}_{\boldsymbol{\sigma}}\Big[\Big|\frac{1}{n}\sum_{i=1}^{n}\sigma_i g(\mathbf{x}_i)\Big|\Big] && \text{(Triangle Inequality)} \\ &= \mathcal{R}_n(\mathcal{F}) + \mathbb{E}_{\boldsymbol{\sigma}}\Big|\frac{1}{n}\sum_{i=1}^{n}\sigma_i g(\mathbf{x}_i)\Big| \\ &\le \mathcal{R}_n(\mathcal{F}) + \frac{1}{n}\sqrt{\mathbb{E}_{\boldsymbol{\sigma}}\Big|\sum_{i=1}^{n}\sigma_i g(\mathbf{x}_i)\Big|^2} && \text{(Jensen's Inequality)} \\ &= \mathcal{R}_n(\mathcal{F}) + \sqrt{\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sum_{i=1}^{n}\sigma_i^2 g^2(\mathbf{x}_i) + \sum_{i\neq j}\sigma_i\sigma_j g(\mathbf{x}_i)g(\mathbf{x}_j)\Big]} \\ &= \mathcal{R}_n(\mathcal{F}) + \sqrt{\mathbb{E}_{\boldsymbol{\sigma}}\sum_{i=1}^{n}\sigma_i^2 g^2(\mathbf{x}_i)} \le \mathcal{R}_n(\mathcal{F}) + \frac{||g||_\infty}{\sqrt{n}}.\end{aligned}$$

Which concludes the proof. □

3 Complexity of Unit Balls

Let $\mathcal{F} = \{x \to \langle \boldsymbol{\theta}, \mathbf{x} \rangle \mid \boldsymbol{\theta} \in \Theta\}$ be a hypothesis class. In this section, we will bound the Radamacher complexity of $\mathcal{F}$ for different choices of Θ.

3.1 L_2 Ball

Assume that $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n \mid ||\theta||_2 \leq r\}$ and let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be the sample points. We can write the following chain of inequalities for the empirical Radamacher complexity.

$$
\begin{aligned}
\mathcal{R}_S(\mathcal{F}) &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{||\theta||_2 \leq r} \frac{1}{n}\sum_{i=1}^n \sigma_i \langle \mathbf{x}_i, \boldsymbol{\theta}\rangle\Big] = \frac{1}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{||\theta||_2 \leq r} \Big\langle \sum_{i=1}^n \sigma_i \mathbf{x}_i, \boldsymbol{\theta}\Big\rangle\Big] \\
&\leq \frac{1}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{||\theta||_2 \leq r} \big|\big|\sum_{i=1}^n \sigma_i \mathbf{x}_i\big|\big|_2 \big|\big|\boldsymbol{\theta}\big|\big|_2\Big] && \text{(Cauchy-Schwartz)} \\
&\leq \frac{r}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\big|\big|\sum_{i=1}^n \sigma_i \mathbf{x}_i\big|\big|_2\Big] = \frac{r}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\big|\big|\sum_{i=1}^n (\sigma_i x_{i1}, \ldots, \sigma_i x_{id})^\top\big|\big|_2\Big] \\
&= \frac{r}{n}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sqrt{\sum_{j=1}^d \Big(\sum_{i=1}^n \sigma_i x_{ij}\Big)^2}\right] \\
&\leq \frac{r}{n}\sqrt{\sum_{j=1}^d \mathbb{E}_{\boldsymbol{\sigma}}\Big(\sum_{i=1}^n \sigma_i x_{ij}\Big)^2} && \text{(Jensen's Inequality)} \\
&\leq \frac{r}{n}\sqrt{d \max_j \left[\mathbb{E}_{\boldsymbol{\sigma}}\Big(\sum_{i=1}^n \sigma_i x_{ij}\Big)^2\right]} \leq \frac{r}{n}\sqrt{dnb^2} = \frac{rb\sqrt{d}}{\sqrt{n}}.
\end{aligned}
$$

3.2 L_1 Ball

Now, consider $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n \mid ||\theta||_1 \leq r\}$. The empirical Radamacher complexity can be bounded from above as follows:

$$
\begin{aligned}
\mathcal{R}_S(\mathcal{F}) &= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{||\theta||_1 \leq r} \frac{1}{n}\sum_{i=1}^n \sigma_i \langle \mathbf{x}_i, \boldsymbol{\theta}\rangle\Big] = \frac{1}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{||\theta||_1 \leq r} \Big\langle \sum_{i=1}^n \sigma_i \mathbf{x}_i, \boldsymbol{\theta}\Big\rangle\Big] \\
&\leq \frac{1}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{||\theta||_1 \leq r} \big|\big|\sum_{i=1}^n \sigma_i \mathbf{x}_i\big|\big|_\infty \big|\big|\boldsymbol{\theta}\big|\big|_1\Big] \leq \frac{r}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\ \big|\big|\sum_{i=1}^n \sigma_i \mathbf{x}_i\big|\big|_\infty\Big] \leq \frac{r}{n}\mathbb{E}_{\boldsymbol{\sigma}}\Big[\ \max_{j\in[d]} \underbrace{\sum_{i=1}^n \sigma_i x_{ij}}_{H_j}\Big],
\end{aligned}
$$

where the for to prove the last inequalty, we used Hölder. The term $\mathbb{E}_{\boldsymbol{\sigma}}\Big[\max_{j\in[d]}(H_j)\Big]$ can be bounded from above by $b\sqrt{2\log(d)}$ using Massart's Lemma [MRT19]:

$$
\begin{aligned}
\exp\Big(\lambda \mathbb{E}\max_{j\in[d]} \sum_{i=1}^n \sigma_i x_{ij}\Big) &\leq \mathbb{E}\Big[\exp\big(\lambda \mathbb{E}\max_{j\in[d]} \sum_{i=1}^n \sigma_i x_{ij}\big)\Big] && \text{(Jensen's Inequality)} \\
&\leq \mathbb{E}\Big[\max_{j\in[d]} \exp\big(\lambda \sum_{i=1}^n \sigma_i x_{ij}\big)\Big] \\
&\leq \mathbb{E}\Big[\sum_{j=1}^d \exp\big(\lambda \sum_{i=1}^n \sigma_i x_{ij}\big)\Big] \\
&\leq \sum_{j=1}^d \prod_{i=1}^n \mathbb{E}e^{\lambda\sigma_i x_{ij}} = \sum_{j=1}^d \prod_{i=1}^n \Big[\frac{1}{2}e^{\lambda x_{ij}} + \frac{1}{2}e^{-\lambda x_{ij}}\Big] \\
&= \sum_{j=1}^d \prod_{i=1}^n e^{\lambda x_{ij}^2/2} \leq d\exp\Big(\frac{\lambda^2 b^2}{2}\Big).
\end{aligned}
$$

By optimizing λ, we can conclude that $\mathbb{E}_{\boldsymbol{\sigma}}\left[\max_{j\in[d]}\sum_{i=1}^{n}\sigma_i x_{ij}\right] \leq b\sqrt{2\log(d)}$. Thus

$$\mathcal{R}_S(\mathcal{F}) \leq \frac{rb\sqrt{2\log(d)}}{\sqrt{n}}.$$

References

[MRT19] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning.* The MIT Press, second edition, 2019.