COS-511: Learning Theory **Spring 2017**

Lecture 12: Rademacher Bounds and Rademacher Calculus

Lecturer: Roi Livni

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

12.1 Rademacher Bounds

Lemma 12.1 (Masart's Lemma). *Let A be some finite set of vectors in $\mathbb{R}^m$ s.t. $\|a\| \le r$ then:*

$$\mathbb{E}\left[\sup_{a\in A}\frac{1}{m}\sum_{i=1}^{m}\sigma_i a_i\right] \le \frac{r\sqrt{2\ln|A|}}{m}$$

Proof. Set

$$\mu = \mathbb{E}\left[\sup_{a\in A}\frac{1}{m}\sum_{i=1}^{m}\sigma_i a_i\right]$$

Then for every $\lambda > 0$

$$\begin{aligned}
e^{\lambda\mu} &\le \mathbb{E}\left[\exp(\lambda\sup_{a\in A}\frac{1}{m}\sum_{i=1}^{m}\sigma_i a_i)\right] \\
&= \mathbb{E}\left[\sup_{a\in A}\exp(\lambda\sum_{i=1}^{m}\sigma_i a_i)\right] \le \mathbb{E}\left[\sum_{a\in A}\exp(\lambda\sum_{i=1}^{m}\sigma_i a_i)\right] = \sum_{a\in A}\mathbb{E}\left[\exp(\lambda\sum_{i=1}^{m}\sigma_i a_i)\right] \\
&= \sum_{a\in A}\prod_{i=1}^{m}\mathbb{E}\left[\exp(\lambda\sigma_i a_i)\right] \le \sum_{a\in A}\prod_{i=1}^{m}\frac{1}{2}\left[\exp(\lambda a_i)+\exp(-\lambda a_i)\right] \\
&\le \sum_{a\in A}\prod_{i=1}^{m}\left[\exp(\lambda a_i^2/2)\right] && : \frac{e^x+e^{-x}}{2} \le e^{x^2/2} \\
&= \sum_{a\in A}\left[\exp(\lambda\|a\|^2/2)\right] \le |A|e^{\lambda r^2/2}
\end{aligned}$$

Taking log and dividng by λ we get that:

$$\mu \le \frac{\ln|A|}{\lambda} + \frac{\lambda r^2}{2}$$

Taking $\lambda = r\sqrt{2\ln|A|}$ we obtain the Lemma. ∎

Corollary 12.2 (Finite Classes). *Let $\mathcal{F}$ be a finite set of functions such that $|f(\mathbf{z})| \le 1$ then:*

$$\mathfrak{R}_m(\mathcal{F}) \le \sqrt{\frac{2\ln|F|}{m}}$$

Proof. Given $\mathcal{F}$ and a sample $S = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ consider the set of vectors $A = \{(f(\mathbf{z}_1), \ldots, f(\mathbf{z}_m)) : f \in \mathcal{F}\}$, then note that for every $a \in A$ we have $\|a\| = \sqrt{\sum_{i=1}^m f(\mathbf{z}_i)} \le \sqrt{m}$. Applying Massart's Lemma we obtain that

$$\mathfrak{R}_S(\mathcal{F}) = \mathbf{E}\left[\sup_{a \in A} \left|\sum \sigma_i a_i\right|\right] \le \sqrt{\frac{2 \ln |\mathcal{F}|}{m}}.$$

□

Lemma 12.3. *Let H be a Hilbert space (for simplicity assume $H = \mathbb{R}^d$) and define $\mathcal{F} = \{f_{\mathbf{w}}(\mathbf{x}) : f_{\mathbf{w}}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle, \|\mathbf{w}\| \le 1\}$ Then*

$$\mathfrak{R}_S(\mathcal{F}) \le \frac{\max_i \|\mathbf{x}^{(i)}\|_2}{\sqrt{m}}$$

Proof.

$$\begin{aligned}
m\mathfrak{R}_S(\mathcal{F}) &= \mathbb{E}\left[\sup_{\|\mathbf{w}\| \le 1} \sum_{i=1}^m \sigma_i \langle \mathbf{w}, \mathbf{x}^{(i)} \rangle\right] \\
&= \mathbb{E}\left[\sup_{\|\mathbf{w}\| \le 1} \langle \mathbf{w}, \sum_{i=1}^m \sigma_i \mathbf{x}^{(i)} \rangle\right] \\
&= \mathbb{E}\left[\left\|\sum_{i=1}^m \sigma_i \mathbf{x}^{(i)}\right\|\right] = \mathbb{E}\left[\sqrt{\left\|\sum_{i=1}^m \sigma_i \mathbf{x}^{(i)}\right\|^2}\right] \\
&\le \sqrt{\mathbb{E}\left[\left\|\sum_{i=1}^m \sigma_i \mathbf{x}^{(i)}\right\|^2\right]} && \text{Concavity of } \sqrt{} \\
&= \sqrt{\mathbb{E}\left[\sum_{i,j=1}^m \sigma_i \sigma_j \langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \rangle\right]} \\
&= \sqrt{\mathbb{E}\left[\sum_{i=1}^m \|\mathbf{x}^{(i)}\|^2\right]} && \mathbb{E}(\sigma_i \sigma_j) = \begin{cases} 1 & i = j \\ 0 & \text{o.w.} \end{cases} \\
&\le \sqrt{m \max \|\mathbf{x}^{(i)}\|^2}
\end{aligned}$$

□

12.1.1 Rademacher Calculus

The following facts are easy to prove, and we leave them as an exercise:

Fact 12.1.

1. *Let $c \cdot \mathcal{F} + b = \{c \cdot f + b : f \in \mathcal{F}\}$. Then*

$$\mathfrak{R}(c \cdot \mathcal{F} + b) = |c| \mathfrak{R}(\mathcal{F})$$

2. *Let* $\text{conv}\mathcal{F} = \{\sum \alpha_i f_i : \{f_i\} \subseteq \mathcal{F}\ \alpha_i \geq 0,\ \sum \alpha_i = 1\}$. *Then*

$$\mathfrak{R}(\text{conv}\mathcal{F}) = \mathfrak{R}(\mathcal{F})$$

Lemma 12.4. *Let* $\phi_{\mathbf{z}}$ *be* ρ*-Lipschitz functions for every* $\mathbf{z} \in \mathcal{X}$ *(i.e.* $|\phi_{\mathbf{z}}(a) - \phi_{\mathbf{z}}(b)| \leq \rho|a-b|$*). Denote*

$$\phi \circ \mathcal{F} = \{\phi_{\mathbf{z}}(f(\mathbf{z})) : f \in \mathcal{F}\}.$$

Then

$$\mathfrak{R}_m(\phi \circ \mathcal{F}) \leq \rho \mathfrak{R}_m(\mathcal{F})$$

Proof. w.l.o.g we may assume $\rho = 1$, the more general case will follow by settin $\phi'_{\mathbf{z}} = \frac{1}{\rho}\phi$ and applying property 1. Given $S = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$, Let us write

$$\mathfrak{R}_S(\phi_{[t]} \circ \mathcal{F}) = \mathop{\mathbb{E}}_{\sigma}\left[\left|\sup_{f\in\mathcal{F}} \sum_{i\leq t} \sigma_i \phi_{\mathbf{z}_i}(f(\mathbf{z}_i)) + \sum_{i>t} \sigma_i f(\mathbf{z}_i)\right|\right]$$

We will show by induction that $\mathfrak{R}_S(\phi_{[t]} \circ \mathcal{F}) \leq \mathfrak{R}_S(\mathcal{F})$. Since the case for $t = 1$ is similar to the induction step, we will show the proof only for $t = 1$. Let us write $\pm\mathcal{F} = \{\pm f : f \in \mathcal{F}\}$ then:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma}\left[\left|\sup_{f\in\mathcal{F}} \sigma_1 \phi_{\mathbf{z}_1}(f(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f(\mathbf{z}_i)\right|\right] \\
=&\frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\left|\sup_{f\in\mathcal{F}} \phi_{\mathbf{z}_1}(f(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f(\mathbf{z}_i)\right|\right] + \frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\left|\sup_{f'\in\mathcal{F}} -\phi_{\mathbf{z}_1}(f'(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right|\right] \\
=& \frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\left|\sup_{f\in\mathcal{F}} \phi_{\mathbf{z}_1}(f(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f(\mathbf{z}_i)\right| + \left|\sup_{f'\in\mathcal{F}} -\phi_{\mathbf{z}_1}(f'(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right|\right] \\
=& \frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{f\in\pm\mathcal{F}}\left[\phi_{\mathbf{z}_1}(f(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f(\mathbf{z}_i)\right] + \sup_{f'\in\pm\mathcal{F}}\left[-\phi_{\mathbf{z}_1}(f'(\mathbf{z}_1)) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right]\right] \\
=& \frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{f,f'\in\pm\mathcal{F}} (\phi_{\mathbf{z}_1}(f(\mathbf{z}_1)) - \phi_{\mathbf{z}_1}(f'(\mathbf{z}_1))) + \sum_{i>1} \sigma_i f(\mathbf{z}_i) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right] \\
\leq& \frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{f,f'\in\pm\mathcal{F}} |f(\mathbf{z}_1) - f'(\mathbf{z}_1)| + \sum_{i>1} \sigma_i f(\mathbf{z}_i) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right]
\end{aligned}$$

Next, we claim that we can remove the absolute value from the term $|f(\mathbf{z}_1) - f'(\mathbf{z}_1)|$, since we are taking supremum over the terms, and the sum of the two other terms will not be effected by replacing f with f', hence:

$$\begin{aligned}
&\frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{f,f'\in\pm\mathcal{F}} |f(\mathbf{z}_1) - f'(\mathbf{z}_1)| + \sum_{i>1} \sigma_i f(\mathbf{z}_i) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right] \\
&= \frac{1}{2}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{f,f'\in\pm\mathcal{F}} f(\mathbf{z}_1) - f'(\mathbf{z}_1) + \sum_{i>1} \sigma_i f(\mathbf{z}_i) + \sum_{i>1} \sigma_i f'(\mathbf{z}_i)\right] = \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\sup_{f,f'\in\pm\mathcal{F}} \sigma_1 f(\mathbf{z}_1) + \sum_{i>1} \sigma_i f(\mathbf{z}_i)\right] = \mathfrak{R}_S(\mathcal{F})
\end{aligned}$$

□