**CMSC 35900 (Spring 2008) Learning Theory** **Lecture: 10**

# Massart's Finite Class Lemma and Growth Function

*Instructors: Sham Kakade and Ambuj Tewari*

## 1 Growth function

Consider the case $\mathcal{Y} = \{\pm 1\}$ (classification). Let $\phi$ be the 0-1 loss function and $\mathcal{F}$ be a class of $\pm 1$-valued functions. We can relate the Rademacher average of $\phi_{\mathcal{F}}$ to that of $\mathcal{F}$ as follows.

**Lemma 1.1.** *Suppose $\mathcal{F} \subseteq \{\pm 1\}^{\mathcal{X}}$ and let $\phi(y', y) = \mathbf{1}[y' \neq y]$ be the 0-1 loss function. Then we have,*

$$\mathfrak{R}_m(\phi_{\mathcal{F}}) = \frac{1}{2}\mathfrak{R}_m(\mathcal{F}) \ .$$

*Proof.* Note that we can write $\phi(y', y)$ as $(1 - yy')/2$. Then we have,

$$\begin{aligned}
\mathfrak{R}_m(\phi_{\mathcal{F}}) &= \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{m}\sum_{i=1}^m \epsilon_i \frac{1 - Y_i f(X_i)}{2} \middle| X_1^m, Y_1^m\right] \\
&= \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{m}\sum_{i=1}^m \epsilon_i \frac{Y_i f(X_i)}{2} \middle| X_1^m, Y_1^m\right] \qquad (1) \\
&= \frac{1}{2}\mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{m}\sum_{i=1}^m (-\epsilon_i Y_i) f(X_i) \middle| X_1^m, Y_1^m\right] \\
&= \frac{1}{2}\mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{m}\sum_{i=1}^m \epsilon_i f(X_i) \middle| X_1^m, Y_1^m\right] \qquad (2) \\
&= \frac{1}{2}\mathfrak{R}_m(\mathcal{F}) \ .
\end{aligned}$$

Equation (1) follows because $\mathbb{E}\left[\epsilon_i | X_1^m, Y_1^m\right] = 0$. Equation (2) follows because $-\epsilon_i Y_i$'s jointly have the same distribution as $\epsilon_i$'s. □

Note that the Rademacher average of the class $\mathcal{F}$ can also be written as

$$\mathfrak{R}_m(\mathcal{F}) = \mathbb{E}\left[\sup_{a \in \mathcal{F}_{|X_1^m}} \frac{1}{m}\sum_{i=1}^m \epsilon_i a_i\right] \ ,$$

where $\mathcal{F}_{|X_1^m}$ is the function class $\mathcal{F}$ restricted to the set $X_1, \ldots, X_m$. That is,

$$\mathcal{F}_{|X_1^m} := \{((f(X_1), \ldots, f(X_m)) \mid f \in \mathcal{F}\} \ .$$

Note that $\mathcal{F}_{|X_1^m}$ is finite and

$$|\mathcal{F}_{|X_1^m}| \leq \min\{|\mathcal{F}|, 2^m\} \ .$$

Thus we can define the *growth function* as

$$\Pi_{\mathcal{F}}(m) := \max_{x_1^m \in \mathcal{X}^m} |\mathcal{F}_{|x_1^m}| \ .$$

The following lemma due to Massart allows us to bound the Rademacher average in terms of the growth function.

**Finite Class Lemma (Massart).** *Let $\mathcal{A}$ be some finite subset of $\mathbb{R}^m$ and $\epsilon_1, \ldots, \epsilon_m$ be independent Rademacher random variables. Let $r = \sup_{a \in \mathcal{A}} \|a\|$. Then, we have,*

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

*Proof.* Let

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

We have, for any $\lambda > 0$,

$$\begin{aligned}
e^{\lambda\mu} &\leq \mathbb{E}\left[\exp\left(\lambda \sup_{a \in \mathcal{A}} \sum_{i=1}^{m} \epsilon_i a_i\right)\right] && \text{Jensen's inequality} \\
&= \mathbb{E}\left[\sup_{a \in \mathcal{A}} \exp\left(\lambda \sum_{i=1}^{m} \epsilon_i a_i\right)\right] \\
&\leq \mathbb{E}\left[\sum_{a \in \mathcal{A}} \exp\left(\lambda \sum_{i=1}^{m} \epsilon_i a_i\right)\right] \\
&= \sum_{a \in \mathcal{A}} \mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{m} \epsilon_i a_i\right)\right] \\
&= \sum_{a \in \mathcal{A}} \prod_{i=1}^{m} \mathbb{E}\left[\exp\left(\lambda \epsilon_i a_i\right)\right] \\
&= \sum_{a \in \mathcal{A}} \prod_{i=1}^{m} \frac{e^{\lambda a_i} + e^{-\lambda a_i}}{2} \\
&\leq \sum_{a \in \mathcal{A}} \prod_{i=1}^{m} e^{\lambda^2 a_i^2/2} && \because \frac{e^x + e^{-x}}{2} \leq e^{x^2/2} \\
&= \sum_{a \in \mathcal{A}} e^{\lambda^2 \|a\|^2/2} \\
&\leq |\mathcal{A}| e^{\lambda^2 r^2/2}
\end{aligned}$$

Taking logs and dividing by $\lambda$, we get that, for any $\lambda > 0$,

$$\mu \leq \frac{\ln |\mathcal{A}|}{\lambda} + \frac{\lambda r^2}{2} \ .$$

Setting $\lambda = \sqrt{2 \ln |\mathcal{A}| / r^2}$ gives,

$$\mu \leq r\sqrt{2 \ln |\mathcal{A}|} \ ,$$

which proves the lemma. □