Massart’s Finite Class Lemma and Growth Function

1 Growth function

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

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

$$R_m(\phi_F) = \frac{1}{2} R_m(F).$$

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

$$R_m(\phi_F) = \mathbb{E} \left[ \sup_{f \in F} \frac{1}{m} \sum_{i=1}^{m} \epsilon_i \left(1 - \frac{Y_i f(X_i)}{2}\right) X_i^m, Y_i^m \right]
= \mathbb{E} \left[ \sup_{f \in F} \frac{1}{m} \sum_{i=1}^{m} \epsilon_i Y_i f(X_i) \left| X_i^m, Y_i^m \right| \right]
= \frac{1}{2} \mathbb{E} \left[ \sup_{f \in F} \frac{1}{m} \sum_{i=1}^{m} \epsilon_i f(X_i) \left| X_i^m, Y_i^m \right| \right]
= \frac{1}{2} \mathbb{E} \left[ \sup_{f \in F} \frac{1}{m} \sum_{i=1}^{m} \epsilon_i f(X_i) \right| X_i^m, Y_i^m \right]
= \frac{1}{2} R_m(F).$$

Equation (1) follows because $\mathbb{E} [\epsilon_i | X_i^m, Y_i^m] = 0$. 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 $F$ can also be written as

$$R_m(F) = \mathbb{E} \left[ \sup_{a \in F_1^m} \frac{1}{m} \sum_{i=1}^{m} \epsilon_i a_i \right],$$

where $F_1^m$ is the function class $F$ restricted to the set $X_1, \ldots, X_m$. That is,

$$F_1^m := \{(f(X_1), \ldots, f(X_m)) | f \in F\}.$$

Note that $F_1^m$ is finite and

$$|F_1^m| \leq \min\{|F|, 2^m\}.$$

Thus we can define the *growth function* as

$$\Pi_F(m) := \max_{x^m \in X^m} |F_1^m|.$$

The following lemma due to Massart allows us to bound the Rademacher average in terms of the growth function.
**Finite Class Lemma (Massart).** Let \( A \) be some finite subset of \( \mathbb{R}^m \) and \( \epsilon_1, \ldots, \epsilon_m \) be independent Rademacher random variables. Let \( r = \sup_{a \in A} \|a\| \). Then, we have,

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

**Proof.** Let

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

We have, for any \( \lambda > 0 \),

\[
e^{\lambda \mu} \leq E \left[ \exp \left( \lambda \sup_{a \in A} \sum_{i=1}^{m} \epsilon_i a_i \right) \right]\quad \text{Jensen’s inequality}
= E \left[ \sup_{a \in A} \exp \left( \lambda \sum_{i=1}^{m} \epsilon_i a_i \right) \right]
\leq E \left[ \sum_{a \in A} \exp \left( \lambda \sum_{i=1}^{m} \epsilon_i a_i \right) \right]
= \sum_{a \in A} E \left[ \exp \left( \lambda \sum_{i=1}^{m} \epsilon_i a_i \right) \right]
= \sum_{a \in A} \prod_{i=1}^{m} E \left[ \exp (\lambda \epsilon_i a_i) \right]
= \sum_{a \in A} \prod_{i=1}^{m} e^{\lambda a_i} + e^{-\lambda a_i} / 2
\leq \sum_{a \in A} \prod_{i=1}^{m} e^{\lambda^2 a_i^2 / 2}
\leq \sum_{a \in A} e^{\lambda^2 \|a\|^2 / 2}
\leq |A| e^{\lambda^2 r^2 / 2}
\]

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

\[
\mu \leq \frac{\ln |A|}{\lambda} + \frac{\lambda r^2}{2}.
\]

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

\[
\mu \leq r \sqrt{2 \ln |A|},
\]

which proves the lemma. \( \square \)