CS 7545: Machine Learning Theory **Fall 2019**

Lecture 21: Massart's Lemma and Sauer's Lemma

Lecturer: Jacob Abernethy *Scribes: Dongliang Zheng, Hao Zhou*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

21.1 Review

- Let G be a class of functions: $z \to [0,1]$.
- Let $s \subseteq Z, s = \{x_1, x_2, ..., x_n\}$.
- Let $\hat{E}_s g = \frac{1}{m}\sum_{i=1}^{m} g(x_i)$ donate empirical mean, and $Eg = E_{Z\sim D[g(Z)]}$ true mean. Where distribution $D \in \Delta(Z)$.
- Hoeffding: fix function g, assume a set of samples $s \sim D$ i.i.d. Then $\hat{E}_s g - Eg \leq \sqrt{\frac{log1/\delta}{m}}$, with probability $\geq 1-\delta$.
- Now, want to find a upper bound for $sup_{g\in G}\hat{E}_s g - Eg$, for $\forall g \in G$.

Last time showed $\Phi(s) \leq 2\mathcal{R}_m(G) + \sqrt{\frac{log1/\delta}{m}}$, where $\mathcal{R}_m(G) = E_{s\sim D}E_{\sigma_1,\sigma_2,...,\sigma_m}[sup_{g\in G}\frac{1}{m}\sum_{i=1}^{m} g(x_i)\delta_i]$.

Given $s \subseteq Z$, $s = \{x_1, x_2, ..., x_n\}$, write $G|_s$ as $G|_s = \{(g(x_1), g(x_2), ..., g(x_n)) : g \in G\}$.

Recall the growth function $\Pi_G(m) = max_{s\subseteq Z,|s|=m}|G|_s|$, where G is assumed to be binary class of function.

21.2 Massart's Lemma

Lemma 21.1 *Let $A \subseteq R^m$, a finite set, and $max_{a\in A}\|a\|_2 \leq r$, then*

$$E_{\sigma_1,\sigma_2,...,\sigma_m}[sup_{a\in A}\sum_{i=1}^{m} a_i\sigma_i] \leq r\sqrt{log2|A|}$$

.

(Note that the left term is Rademacher complexity-like and the right side is bounded by its size)

Proof: For any $\lambda > 0$,

$$\begin{aligned}
exp(\lambda E_{\sigma_1,\ldots,\sigma_m}[sup_{a\in A} a\sigma]) &\leq E_{\sigma_1,\ldots,\sigma_m}[exp(\lambda sup_{a\in A} a\sigma]) \\
&\leq E_{\sigma_1,\ldots,\sigma_m}[sup_{a\in A} exp(\lambda a\sigma] \\
&\leq E_{\sigma_1,\ldots,\sigma_m}[\sum_{a\in A} exp(\lambda a\sigma)] \\
&= E_{\sigma_1,\ldots,\sigma_m}[\sum_{a\in A} exp(\lambda \sum_{i=1}^{m}(a_i\sigma_i))] \\
&= \sum_{a\in A} E_{\sigma_1,\ldots,\sigma_m}[\prod_{i=1}^{m} exp(\lambda a_i\sigma_i)] \\
&= \sum_{a\in A}\prod_{i=1}^{m} E_{\sigma_1,\ldots,\sigma_m}[exp(\lambda a_i\sigma_i)] \\
&\leq \sum_{a\in A}\prod_{i=1}^{m} exp\left(\frac{\lambda^2(2a_i)^2}{8}\right) \quad (Hoeffding's\ lemma) \\
&= \sum_{a\in A} exp(\frac{\lambda^2}{2}\sum_{i=1}^{m} a_i^2) \\
&\leq |A| exp(\frac{\lambda^2}{2} r^2)
\end{aligned}$$

Take *log* of both side of the above inequality, and multiply by $1/\lambda$, we get

$$E_{\sigma_1,\ldots,\sigma_m}[sup_{a\in A} a\sigma] \leq \frac{log|A|}{\lambda} + \frac{r^2\lambda}{2}$$

Set $\lambda = \sqrt{\frac{2log|A|}{r^2}}$, we get the inequality in the lemma. ■

Notice that:

$$\begin{aligned}
\hat{R}_S(G) :&= E_{\sigma_{1:m}}[\frac{1}{m}\sup_{g\in G}\Sigma_{i=0}^{m} g(x_i)\sigma_i] \\
&= E_{\sigma_{1:m}}[\frac{1}{m}\sup_{a\in G|s} a\sigma] \\
&\leq \frac{1}{m}\sqrt{m}\sqrt{2log|G|s|} \\
&\leq \sqrt{\frac{2log\Pi_G(m)}{m}}
\end{aligned}$$

note that $\sqrt{m}$ is the r in lemma.

21.3 Sauer's Lemma

Lemma 21.2 *If binary function class G has VC-dim=d, then:* $\Pi_G(m) \leq \Sigma_{i=1}^{d}\binom{m}{i} \leq m^d$

Proof: Let $M_{S,G}$ be the matrix whose unique rows are $\{g(x_1), g(x_2)...g(x_m)\}$, delete repeated rows.
Fact: $\Pi_G(m) = \max_{S\subseteq R, |S|=m} \#rows(M_{S,G})$
Fact: if all rows of $M_{S,G}$ had $\leq d$ $\mathbb{1}$'s, then:# rows($M_{S,G}$) $\leq \Sigma_{i=0}^{d}\binom{m}{i}$
Trick: modify $M_{S,G}$ so that every row has $\leq$ d $\mathbb{1}$s and vc-dim does not increase, and no row has duplicates

in process.
Procedure: change the $\mathbb{1}$ to 0 in matrix $M_{S,G}$ ■
Proof continued in next lecture.