CMPUT 654 Fa 23: Theoretical Foundations of Machine Learning **Fall 2023**

Lecture 11: Empirical Covering Number Bounds (October 10)

Lecturer: Csaba Szepesvári *Scribes: Kushagra Chandak*

Note: *LaTeX template courtesy of UC Berkeley EECS dept. (link to directory)*
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.*

Now we state the one-sided uniform deviation bounds (similar to additive and multiplicative Chernoff) with L_1 uniform covering number:

Theorem 11.1. *Let* $P \in \mathcal{M}_1(\mathcal{Z})$, $\mathcal{G} \subseteq [0,1]^{\mathcal{Z}}$, *and* $\delta \in (0,1)$. *Then*

1. *w.p.* $1-\delta$, $\forall g \in \mathcal{G}$: $Pg \le P_n g + \inf_{\varepsilon>0}\left(2\varepsilon + 3\sqrt{\frac{\ln(2N_1(\varepsilon,\mathcal{G},n)+1)/\delta}{2n}}\right)$.
2. $\forall z \in (0,1)$ *w.p.* $1-\delta$, $\forall g \in \mathcal{G}$: $(1-z^2)Pg \le P_n g + \inf_{\varepsilon>0}\left(2\varepsilon + \frac{5-4z}{z}\frac{\ln(2N_1(\varepsilon,\mathcal{G},n)+1)/\delta}{2n}\right)$.

Proof.

Lemma 11.2. $\forall a \in \mathbb{R}^n$, $\sigma \sim \mathrm{Rad}(n)$. *Then w.p.* $1-\delta$

$$\frac{1}{n}\langle a, \sigma\rangle \le \sqrt{\frac{2\|a\|_2^2 \ln(1/\delta)}{n^2}}\,.$$

□

For the second part of the theorem, $z \in (0,1)$ is a free parameter which can be chosen as an instance-dependent parameter (e.g. true loss) in applications to get small loss bounds. It is of the order of $1/\sqrt{n}$ which gives a $1/\sqrt{n}$ rate.

Remark 11.3. Lemma 11.2 holds for random functions in the cover. To make the bound hold for uniformly (for all functions in the function class), we average it using σ's. To get rid of the σ's from the final result, we use **??** and **??**. verify