Controlling $\sup_{\theta \in \Theta} |R_n(\hat{g}_\theta) - R(g_\theta)|$

<u>def</u> (Empirical Rademacher Complexity)

$$\hat{\mathcal{E}}_R = \mathbb{E}_\sigma \left[\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i \, l_i(g_\theta) \right| \right] \quad \text{w/ } \sigma_i \overset{iid}{\sim} \text{unif}\{-1, 1\}$$

Rademacher R.V.

$$\underline{\mathbb{E} \sup_{\theta \in \Theta} |R_n(g_\theta) - R(g_\theta)|} = \mathbb{E} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} (l_i - \mathbb{E} l_i)(g_\theta) \right|$$

$$= \mathbb{E} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{l'} (l_i - l_i')(g_\theta) \right| \quad \text{w/ } l_i' \overset{iid}{\sim} l_i$$

$$\leq \mathbb{E}\, \mathbb{E}_{l'} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} (l_i - l_i')(g_\theta) \right| \quad \text{Jensen's ineq.}$$

$$= \mathbb{E}\, \mathbb{E}_{l'}\, \mathbb{E}_\sigma \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i (l_i - l_i')(g_\theta) \right| \quad \text{symmetry}$$

$$\leq \overset{\text{w.r.t. everything}}{\mathbb{E}} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i l_i(g_\theta) \right| + \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i l_i'(g_\theta) \right|$$

$$= \underline{2\, \mathbb{E}\, \hat{\mathcal{E}}_R}$$

<u>thm</u> "Excess Risk is bounded by how well learner explains noise" (l_i iid)

$$\mathbb{E} \sup_{\theta \in \Theta} |R_n(g_\theta) - R(g_\theta)| \leq 2\, \mathbb{E}\, \hat{\mathcal{E}}_R$$