Dropping the Realizability assumption:

First we recall a Chernoff bound due to Hoeffding:

<u>Thm</u> Let $X_1, \cdots X_m$ be ind. random vars with X_i taking values in $[a_i, b_i]$, $i \in [m]$. Then for any $\epsilon > 0$ we have:

let $S_m = \sum_{i=1}^{m} X_i$

$$\mathbb{P}\left[[S_m - \mathbb{E}[S_m]] \geq \epsilon\right] \leq e^{-2\epsilon^2 / \sum_{i=1}^{m}(b_i - a_i)^2}$$

$$\mathbb{P}\left[[S_m - \mathbb{E}[S_m]] \leq \epsilon\right] \leq e^{-2\epsilon^2 / \sum_{i=1}^{m}(b_i - a_i)^2}$$

<u>Cor 1</u> Fix $\epsilon > 0$ and let S denote an iid sample of size m. Then $\forall\, h: \mathcal{X} \to \underbrace{\{0,1\}}_{\mathcal{Y}}$ we have the following inequalities

$$\mathbb{P}_{S \sim \mathcal{D}^m}\left[L_S(h) - L(h) \geq \epsilon\right] \leq \exp(-2m\epsilon^2) \quad \text{where } m = |S|$$

$$\mathbb{P}_{S \sim \mathcal{D}^m}\left[L_S(h) - L(h) \leq -\epsilon\right] \leq \exp(-2m\epsilon^2)$$

thus $\mathbb{P}_{S \sim \mathcal{D}^m}\left[|L_S(h) - L(h)| \geq \epsilon\right] \leq 2\exp(-2m\epsilon^2)$.

<u>Proof</u> The samples are drawn iid according to $\mathcal{D}$ & the labels are 0 or 1 so $(b_i - a_i) = 1$. & $\sum_{i=1}^{m}(b_i - a_i)^2 = m$.

Define X_i to be the i^{th} element in the sample & define

$R_i := |h(X_i) - y_i|$ R_i takes values in $\{0,1\}$.

$\sum_{i=1}^{m} R_i = m \cdot L_S(h)$. Recall $\mathbb{E}[L_S(h)] = L(h)$

$L_S(h) - L(h) \geq \epsilon$ is thus $\Leftrightarrow$ $\sum R_i - \mathbb{E}[\sum R_i] \geq m\epsilon$

Putting this into the theorem we get the corollary.

<u>Cor 2.</u> Fix $h : \mathcal{X} \to \{0,1\}$. Then for any $\delta > 0$ we have

$$L(h) \leq L_S(h) + \sqrt{\frac{\log 2/\delta}{2m}}$$

with probability at least $(1-\delta)$.

<u>Proof</u> If we set the RHS of Cor 1 to δ and solve for ϵ we get the result. ■

<u>Example</u> (Mohri) Take a biased coin that lands with H with probability p. Take h to be : "always predict H".

$$L(h) = 1-p \quad \& \quad L_S(h) = q \text{ where } q \text{ depends on } S.$$

We get from Cor 2: $|(1-p)-q| \leq \sqrt{\frac{\log 2/\delta}{2m}}$

Choose $\delta = .02$ & $m = 500$

$$|(1-p)-q| \leq \sqrt{\frac{\cancel{10\log 2}\ 2\log 10}{1000}} \approx .06$$

So $L_S(h)$ may not be small but for a fixed h it gives a good approximation of L.

<u>Thm</u> <u>Learning Bound</u> Let $\mathcal{H}$ be a finite hypothesis set. Then $\forall \delta > 0$, with prob. at least $1-\delta$ we have

$$\forall h \in \mathcal{H} \quad L(h) \leq L_S(h) + \sqrt{\frac{\log|\mathcal{H}| + \log 2/\delta}{2m}}$$

where, as usual, S is drawn iid & has m elements.

<u>Proof</u> $\mathbb{P}[\exists h \in \mathcal{H} \text{ s.t. } |L_S(h) - L(h)| > \epsilon] \leq \sum_{h \in \mathcal{H}} \mathbb{P}[|L_S(h) - L(h)| > \epsilon]$

$$\leq 2|\mathcal{H}| \exp(-2m\epsilon^2) \quad [\text{Using Cor 1}]$$

Set RHS $= \delta$ & get the result. ■