Q: How m ways can $\mathcal{H}$ label a set of m pts? [$\tau_{\mathcal{H}}(m)$ = ?]

Lemma: (Sauer-Shelah-Perles) $\tau_{\mathcal{H}}(m) \le \sum_{i=1}^{VC(\mathcal{H})=d} \binom{m}{i}$

$m \ge d$: $\le \left(\frac{em}{d}\right)^d$ ← Stirling's Formula

$m < d$: $= 2^m$

Pf sketch:

Claim 1: for any set $|S'| \ge d$, $\mathcal{H}$ induces $\le 2^d$ labelings of S'.

Why? If it induces more, $VC(\mathcal{H}) > d$

So, we can ~~count~~ up distinct labelings $\mathcal{H}$ induces on C by (1) taking all subsets of size $\le d$ & assuming it's labeled + and outside $\mathcal{H}_C$ is fixed.

(Formal pf by induction on size of $\mathcal{H}_C$ & shattered sets)

What were we doing again? Oh yeah. Arguing that for $VC(\mathcal{H}) = d$, we get PAC learnability of $\mathcal{H}$, for uniform convergence...

The numerical statement of uniform convergence: For every D, δ, $h \in \mathcal{H}$

$$|L_D(h) - L_S(h)| \le \frac{4 + \sqrt{\log(\tau_{\mathcal{H}}(2m))}}{\delta \cdot \sqrt{m/2}}$$

w.p. $1-\delta$.

Sauer's Lemma

For $m \ge d$

$$\le \frac{4 + \sqrt{\log\left(\frac{2em}{d}\right)^d}}{\delta \sqrt{m/2}}$$

$$\le \frac{4 + \sqrt{d \log\left(\frac{2em}{d}\right)}}{\delta \cdot \sqrt{m/2}}$$

Let's assume $4 < \sqrt{d \log\left(\frac{2em}{d}\right)}$:

$$\leq \frac{2\sqrt{d \log\left(\frac{2em}{d}\right)}}{\delta \cdot \sqrt{m/2}} \leq \epsilon$$ WTS

It suffices to show

$$\frac{4 d \log\left(\frac{2em}{d}\right)}{m/2} \leq (\epsilon\delta)^2$$

$$\Leftarrow \quad \frac{8d}{(\epsilon\delta)^2} \log\left(\frac{2em}{d}\right) \leq m$$

Sufficient: $m \geq \frac{8d}{(\epsilon\delta)^2}\left[4 \log \frac{2e}{d} + 16 \log\left(\frac{2d}{(\delta\epsilon)^2}\right)\right]$

This gives an upper bd on sample complexity of uniform convergence & therefore PAC.

We didn't actually show

For every $D, \delta, h \in \mathcal{H}$

$$|L_D(h) - L_S(h)| \leq \frac{4 + \sqrt{\log(\tau_{\mathcal{H}}(2m))}}{\delta \cdot \sqrt{m/2}}$$

w.p. $1-\delta$.

Claim:

$$\star \quad \mathbb{E}_S\left(\sup_{h \in \mathcal{H}} |L_S(h) - L_D(h)|\right) \leq \frac{4 + \sqrt{\log \tau_{\mathcal{H}}(2m)}}{\sqrt{2m}}$$

+ Markov's inequality

$$\Rightarrow \mathbb{P}\left(\sup_{h \in \mathcal{H}} |L_D(h) - L_S(h)| \geq \epsilon\right) \leq \frac{4 + \sqrt{\lg \tau_{\mathcal{H}}(2m)}}{\epsilon\sqrt{m/2}}$$

$$\epsilon \geq \frac{4 + \sqrt{\lg \tau_{\mathcal{H}}(2m)}}{\sqrt{m/2}\,\delta} \leq \delta$$

★ $\mathbb{E}_S\left[\sup_{h\in\mathcal{H}} \left| L_{\mathcal{D}}(h) - L_S(h)\right|\right] \le \dfrac{4+\sqrt{\log(\tau_{\mathcal{H}}(2m))}}{\sqrt{m/2}}$

Pf $\quad = \mathbb{E}_{S=\{x_i,y_i\}}\left[\sup_{h\in\mathcal{H}} \left|\mathbb{E}_{S'=(x_i',y_i')} L_{S'}(h) - L_S(h)\right|\right]$

Jensen's Inequality
convex f
$\Rightarrow \mathbb{E}[f(z)] \ge f(E(z))$

$\le \mathbb{E}_{S=\{(x_i,y_i)\}}\left[\sup_{h\in\mathcal{H}} \mathbb{E}_{S'=\{(x_i',y_i')\}} \left|L_{S'}(h) - L_S(h)\right|\right]$

$\le \mathbb{E}_{S,S'}\left[\sup_{h\in\mathcal{H}} \left|L_S(h) - L_{S'}(h)\right|\right]$

$= \mathbb{E}_{S,S'}\left[\sup_{h\in\mathcal{H}} \left|\frac{1}{m}\sum_{i=1}^m \ell_h(\underbrace{x_i',y_i'}_{z_i'}) - \ell_h(\underbrace{x_i,y_i}_{z_i})\right|\right]$ ★

$\ell_h(z_i), \ell_h(z_i')$ are iid, so

$\ell_h(z_i) - \ell_h(z_i') \stackrel{\text{dist}}{=} \sigma_i\left(\ell_h(z_i) - \ell_h(z_i')\right)$ $\forall \sigma_i \in \{+1,-1\}$, in particular $\sigma_i = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$

$\Rightarrow$ ★ $\mathbb{E}_{\sigma_1,\dots,\sigma_m}\left[\mathbb{E}_{S,S'}\left[\sup_{h\in\mathcal{H}} \left|\frac{1}{m}\sum \sigma_i\left(\ell_h(z_i) - \ell_h(z_i')\right)\right| \;\Big|\; \sigma\right]\right]$

Linearity of expectation $\Rightarrow \mathbb{E}_{S,S'}\left[\mathbb{E}_\sigma\left[\underline{\qquad\qquad} \;\big|\; S,S'\right]\right]$

S,S' fixed. Only $\mathcal{H}_C$ matters, not $\mathcal{H}$, $C = S\cup S'$, $|C|\le 2m$

$= \mathbb{E}_{S,S'}\left[\mathbb{E}_\sigma \max_{h\in\mathcal{H}_C} \left|\frac{1}{m}\sum \sigma_i\left(\ell_h(z_i) - \ell_h(z_i')\right)\right| S,S'\right]$

$\mathbb{E}[\sigma_i(\ell(h(z_i)) - \ell(h(z_i')))] = 0$

$-1 \le \underline{\qquad} \le 1$

So Hoeffding $|a-b| = 2$

$\mathbb{P}\left[\left|\frac{1}{m}\sum \sigma_i\left[\ell_i(z_i) - \ell_i(z_i')\right]\right| > \rho\right] \le 2e^{-\frac{2m\rho^2}{4}}$

UB $\Rightarrow \mathbb{P}\left[\max_{h\in\mathcal{H}} \qquad\qquad\right] \le 2|\mathcal{H}_C|\, e^{-\frac{2m\rho^2}{4}}$

★★ $\Rightarrow \mathbb{E}\max_{h\in\mathcal{H}_C}\left|\frac{1}{m}\sum \sigma_i\left(\ell_h(z_i) - \ell_h(z_i')\right)\right| \le \dfrac{4+\sqrt{\log \tau_{\mathcal{H}}(2m)}}{\sqrt{m/2}}$

[Lemma A.4 in book]