<u>Recall</u>: $\mathcal{H}_S = \{ h|_S : S \to \{0,1\} : h \in \mathcal{H} \}$

<u>Growth Function</u>

$\tau_{\mathcal{H}} : \mathbb{N} \to \mathbb{N},\ \tau_{\mathcal{H}}(m) := \max_{S \subset X : |S| = m} |\mathcal{H}_S|$

Observation: $VCdim(\mathcal{H}) \geq m$ iff $\tau_{\mathcal{H}}(m) = 2^m$

<u>Lemma (Sauer)</u>

$VCdim(\mathcal{H}) = d < \infty \longrightarrow \tau_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} \binom{m}{i} \quad \forall m \in \mathbb{N}$

<u>Proof</u>

Observation: $|\{B \subset S : \mathcal{H} \text{ shatters } B\}| \leq \sum_{i=0}^{d} \binom{|S|}{i}$

Claim: $|\mathcal{H}_S| \leq |\{B \subset S : \mathcal{H} \text{ shatters } B\}|$

$|S| = 1$: $\emptyset$ vacuously shattered; $|\mathcal{H}_S| = 1 + \begin{cases} 1, & \text{if } S \text{ is shattered} \\ 0, & \text{o/w.} \end{cases}$

$1 < |S|$: Let $S = \{s_1, \ldots, s_m\}$, $S' := S \setminus \{s_m\}$

Define

$\mathcal{H}_0 := \{ h \in \mathcal{H}_{S'} : h \cup (s_m, 0) \in \mathcal{H}_S \vee h \cup (s_m, 1) \in \mathcal{H}_S \} = \mathcal{H}_{S'}$

$\mathcal{H}_1 := \{ h \in \mathcal{H}_{S'} : h \cup (s_m, 0) \in \mathcal{H}_S \wedge h \cup (s_m, 1) \in \mathcal{H}_S \}$

$|\mathcal{H}_S| = |\mathcal{H}_0| + |\mathcal{H}_1|$

$|\mathcal{H}_0| \overset{\text{i.h.}}{\leq} |\{B \subset S' : \mathcal{H} \text{ shatters } B\}| = |\{B \subset S : s_m \notin B \wedge \mathcal{H} \text{ shatters } B\}|$

$\mathcal{H}' := \{ h \cup (s_m, 1) : h \in \mathcal{H}_1 \} \cup \{ h \cup (s_m, 0) : h \in \mathcal{H}_1 \}$

$\mathcal{H}_1 = \mathcal{H}'_{S'}$

$|\mathcal{H}_1| \overset{\text{i.h.}}{\leq} |\{B \subset S' : \mathcal{H}' \text{ shatters } B\}| = |\{B \subset S : s_m \in B \wedge \mathcal{H}' \text{ shatters } B\}|$

$\{B \subset S : \mathcal{H} \text{ shatters } B\} =$

$\quad \{B \subset S : s_m \notin B \wedge \mathcal{H} \text{ shatters } B\} \,\dot\cup\, \{B \subset S : s_m \in B \wedge \mathcal{H} \text{ shatters } B\}$

$\Longrightarrow |\mathcal{H}_S| = |\mathcal{H}_0| + |\mathcal{H}_1| \leq |\{B \subset S : \mathcal{H} \text{ shatters } B\}|$

Theorem:

For every $\delta \in (0,1)$ we have with probability at least $(1-\delta)$ over an i.i.d set $S \sim D$

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

Recap:

S is ε-representative if $\forall h \in \mathcal{H}$

$|L_S(h) - L_D(h)| \leq \varepsilon$

<u>UC</u>: $\forall \varepsilon, \delta > 0,\ |S| \geq m^{UC}_{\mathcal{H}}(\varepsilon, \delta)$, with probability $\geq (1-\delta)$

$\forall h \in \mathcal{H}$: $|L_S(h) - L_D(h)| \leq \varepsilon$.

Competitiveness:

h is (ε, δ)-competitive with $\tilde{h} \in \mathcal{H}$ if wpal $(1-\delta)$

$L_D(h) \leq L_D(\tilde{h}) + \varepsilon$

Nonuniform Convergence:

$\forall \varepsilon, \delta > 0, \tilde{h} \in \mathcal{H}, |S| \geq m^{NUL}_{\mathcal{H}}(\varepsilon, \delta, \tilde{h})$ wpal $(1-\delta)$

$L_D(A(S)) \leq L_D(\tilde{h}) + \varepsilon$

$A(S) \in \mathcal{H}$ output by a nonuniform learner.

<u>Structural Risk Minimisation</u>

Let $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$, where $\mathcal{H}_n$ has UC $\forall n \in \mathbb{N}$

Define $n : \mathcal{H} \to \mathbb{N}$, $n(h) = \min\{n \in \mathbb{N} : h \in \mathcal{H}_n\}$

Fix $m \in \mathbb{N}$. $\varepsilon_n(m, \delta) := \overset{\text{inf ?!}}{\min} \{ \varepsilon \in (0,1) : m^{UC}_{\mathcal{H}_n}(\varepsilon, \delta) \leq m \}$

Hence for any $m \in \mathbb{N}$, $\delta \in (0,1)$ wpal $(1-\delta)$

$\forall h \in \mathcal{H}_n$ $|L_D(h) - L_S(h)| \leq \varepsilon_n(m, \delta)$

Weight function:

$w : \mathbb{N} \to [0,1]$ s.t. $\sum_{n=0}^{\infty} w(n) \leq 1$.

<u>Theorem</u>

wpal $(1-\delta)$ (fixed m)

$\forall n \in \mathbb{N}\ \forall h \in \mathcal{H}_n$: $|L_D(h) - L_S(h)| \leq \varepsilon_n(m, w(n) \cdot \delta)$

In particular:

$\forall h \in \mathcal{H}$: $L_D(h) \leq L_S(h) + \min_{n : h \in \mathcal{H}_n} \varepsilon_n(m, w(n) \cdot \delta)$

<u>Proof</u>

$P(\exists n \in \mathbb{N}\ \exists h \in \mathcal{H}_n : |L_D(h) - L_S(h)| > \varepsilon_n(m, w(n) \cdot \delta))$

$\leq \sum_{n=1}^{\infty} P(\exists h \in \mathcal{H}_n\ (\ldots\ldots))$

$\leq \sum_{n=1}^{\infty} w(n) \cdot \delta \leq \delta$

<u>Corollary</u>

wpal $(1-\delta)$

$\forall h \in \mathcal{H}$: $L_D(h) \leq L_S(h) + \varepsilon_{n(h)}(m, w(n(h)) \cdot \delta)$

SRM: Pick $h \in \operatorname{argmin}_{h \in \mathcal{H}} (L_S(h) + \varepsilon_{n(h)}(m, w(n(h)) \cdot \delta))$

<u>Theorem</u>

Assume $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$, where each $\mathcal{H}_n$ has UC.

for fixed $\tilde{h} \in \mathcal{H}$ and i.i.d sample set S with

$|S| \geq m^{UC}_{\mathcal{H}_{n(\tilde{h})}}(\frac{\varepsilon}{2}, w(n(\tilde{h})) \cdot \delta)$ wpal $(1-\delta)$

$L_D(A(S)) \leq L_D(\tilde{h}) + \varepsilon$

<u>Proof</u>

wpal $(1-\delta)$

$\forall h \in \mathcal{H}$: $L_D(h) \leq L_S(h) + \varepsilon_{n(h)}(m, w(n(h)) \cdot \delta)$ for any fixed $m \in \mathbb{N}$

Now:

$L_D(A(S)) \leq \min_{h \in \mathcal{H}} (L_S(h) + \varepsilon_{n(h)}(m, w(n(h)) \cdot \delta))$

$\leq L_S(\tilde{h}) + \varepsilon_{n(\tilde{h})}(m, w(n(\tilde{h})) \cdot \delta)$

wpal

$L_D(A(S)) \leq L_S(\tilde{h}) + \varepsilon_{n(\tilde{h})}(m^{UC}_{\mathcal{H}_{n(\tilde{h})}}(\frac{\varepsilon}{2}, w(n(\tilde{h})) \cdot \delta), w(n(\tilde{h})) \cdot \delta)$

$\leq L_S(\tilde{h}) + \frac{\varepsilon}{2}$

$L_D(A(S)) \leq L_S(\tilde{h}) + \frac{\varepsilon}{2} \leq (L_D(\tilde{h}) + \frac{\varepsilon}{2}) + \frac{\varepsilon}{2} = L_D(\tilde{h}) + \varepsilon$

<u>Corollary</u>

If $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$, where $\forall n \in \mathbb{N}$ is agnostic PAC-learnable,

then $\mathcal{H}$ is nonuniform learnable

$\mathcal{H}_n := \{ h \in \mathcal{H} : m^{NUL}_{\mathcal{H}}(\frac{1}{8}, \frac{1}{7}, h) \leq n \}$

wpal $\frac{6}{7}$ we have $L_D(\underbrace{A(S)}_{\in \mathcal{H}}) \leq \frac{1}{8}$

$\forall h \in \mathcal{H}_n$

$|S| \geq \cdot\ h = A(S)$

$|S'| \geq n$: $L_D(h) \leq L_D(\tilde{h}) + \frac{1}{8}$

$|S''| \geq n$ $L_D(A(S)) \leq L_D(h) + \frac{1}{8}$

$\forall S \sim D^n : A(S) \in \mathcal{H}_n$

$\forall \tilde{h} \in \mathcal{H}_n$

wpal $(1-\delta)$

$L_D(A(S)) \leq \underset{=0}{L_D(\tilde{h})} + \frac{\varepsilon}{8}$

$|S| \geq n$ $\tilde{h}$