Lower bounds on sample complexity

- Let $H \subseteq Y^X$ be a hypothesis class
- Let $d = VC(H)$
- We want show that

$$m(H, \varepsilon, \delta) \gtrsim \frac{d}{\varepsilon}$$

Theorem: Let H be a hypothesis class over a domain X such that $d = VC(H) \geq 2$. Let $\varepsilon \in (0, 1/8)$. Let $A: (X \times Y)^* \to Y^X$ be any learning algorithm. There exists distribution D over X and $h^* \in H$ such that if S is an i.i.d.

sample from D labeled by h^* then $\text{err}_{D,h^*}(A(S)) > \varepsilon$ with probability at least $\frac{1}{100}$ unless

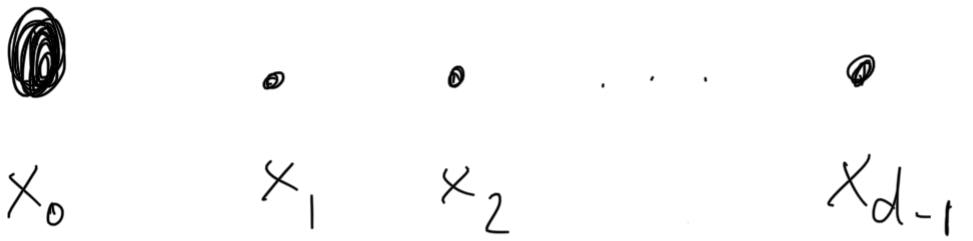
$$m \geq \frac{d-1}{32\varepsilon}.$$

Proof:

- Since $VC(H) = d$, there exists a shattered set of size d.
- Let $C = \{x_0, x_1, x_2 \ldots, x_{d-1}\}$ be a shattered set.
- Define D as follows:

$$\begin{cases} 1 - 8\varepsilon & \text{if } x = x_0 \end{cases}$$

$$D(x) = \begin{cases} \frac{8\varepsilon}{d-1} & \text{if } x \in \{x_1, x_2 \cdots, x_{d-1}\} \\ 0 & \text{if } x \notin C \end{cases}$$



- Let U be an i.i.d. sample from D of size $m < \frac{d-1}{32\varepsilon}$.
- Let Z be the number of samples in $\{x_1, x_2, \cdots, x_{d-1}\}$ (That is, samples other than x_0.)

- $Z = \sum_{i=1}^{m} Z_i$ where $Z_1, Z_2, \ldots, Z_m$ are i.i.d. $\Pr[Z_i = 1] = \delta\varepsilon = p$

- With "high" probability Z is "small".
  We use Chernoff bound:

$$\Pr[Z \geq 2mp] \leq \left(\frac{e}{2^2}\right)^{mp}$$

- $2mp = 2m \cdot \delta\varepsilon \leq 2 \cdot \frac{d-1}{32\varepsilon} \cdot \delta\varepsilon = \frac{d-1}{2}$

- $\left(\frac{e}{2^2}\right)^{mp} = \left(\frac{\sqrt{e}}{2}\right)^{2mp} \leq \left(\frac{\sqrt{e}}{2}\right)^{\frac{d-1}{2}} \leq \left(\frac{\sqrt{e}}{2}\right)^{\frac{1}{2}} \leq \frac{93}{100}$

- So with probability more than

$\frac{7}{100}$, $z < \frac{d-1}{2}$.

- The rest of the proof is similar to the no free lunch theorem.
- Since C is shattered, there exists $2^{|C|}$ functions $h_1, h_2, \ldots, h_{2^{|C|}} \in H$ realizing all possible behaviors on C.
- Let S_i be U labeled by h_i
- Let $T = \{x \in C : x \notin U, x \neq x_0\}$
- $|T| \geq (d-1) - z$

- Condition on the event

$$|T| > \frac{d-1}{2}$$

- $\Pr\left[|T| > \frac{d-1}{2}\right] \geq 1 - \frac{93}{100} = \frac{7}{100}$.

$$\frac{1}{2^{|C|}} \sum_{i=1}^{2^{|C|}} \mathbb{E}\left[\mathrm{err}_{D,h_i}(A(S_i)) \,\middle|\, T > \frac{d-1}{2} \right]$$

$$\geq \frac{1}{2^{|C|}} \sum_{i=1}^{2^{|C|}} \mathbb{E}\left[\frac{8\varepsilon}{d-1} \sum_{x \in T} \mathbb{1}\left[A(S_i)(x) \neq h_i(x)\right] \,\middle|\, T > \frac{d-1}{2} \right]$$

$$= \frac{8\varepsilon}{d-1} \mathbb{E}\left[\mathbb{E}\left[\sum_{x \in T} \frac{1}{2^{|C|}} \sum_{i=1}^{2^{|C|}} \mathbb{1}\left[A(S_i)(x) \neq h_i(x)\right] \,\middle|\, T \right] \,\middle|\, T > \frac{d-1}{2} \right]$$

$$\geq \frac{8\varepsilon}{} \mathbb{E}\left[\frac{|T|}{} \,\middle|\, T > \frac{d-1}{} \right]$$

$$= d-1 \left\lfloor \frac{1}{2} \right\rfloor \qquad \left(\right)$$

$$> \frac{8\varepsilon}{d-1} \cdot \frac{d-1}{4}$$

$$= 2\varepsilon$$

So there exists $h^* \in H$

$$E\left[err_{h^*,D}(A(S)) \;\middle|\; |T| > \frac{d-1}{2} \right] \geq 2\varepsilon$$

Let $Y = 1 - err_{h^*,D}(A(S))$

$$P_r\left[err_{h^*,D}(A(S)) \leq \varepsilon \;\middle|\; |T| > \frac{d-1}{2} \right]$$

$$= \Pr\left[Y \geq 1 - \varepsilon \;\middle|\; |T| > \frac{d-1}{2} \right]$$

$$\leq \frac{\mathbb{E}\left[Y \;\middle|\; |T| > \frac{d-1}{2} \right]}{1 - \varepsilon}$$

$$\leq \frac{1 - 2\varepsilon}{1 - \varepsilon}$$

$$\leq \frac{1 - 2 \cdot 1/8}{1 - 1/8}$$

$\varepsilon \leq \frac{1}{8}$

$$= \frac{3/4}{7/8}$$

6

$= \frac{1}{7}$

- $\Pr\left[err_{D,h^*}(A(S)) > \varepsilon \;\middle|\; |T| > \frac{d-1}{2} \right] \geq \frac{1}{7}$

- $\Pr\left[err_{D,h^*}(A(S)) > \varepsilon \right]$

$\geq \Pr\left[|T| > \frac{d-1}{2} \right] \cdot \Pr\left[err_{D,h^*}(A(S)) \;\middle|\; |T| > \frac{d-1}{2} \right]$

$\geq \frac{7}{100} \cdot \frac{1}{7}$

$= \frac{1}{100}$