Rademacher Complexity and VC dimension Sep / 12 / 2023

§1 Motivation

(1) What can we say if $\mathcal{H}$ is infinite?

(2) Can we measure the complexity of $\mathcal{H}$?

(3) Desired property on the measure.

- Uniform bound on $R, \hat{R}_S$.

(cf) $\underset{S\sim D^m}{\mathbb{P}}\left[\forall h\in\mathcal{H}\ \ R(h)\le \hat{R}_S(h) + \sqrt{\frac{\log|\mathcal{H}|+\log\frac{1}{\delta}}{2m}}\right] \ge 1-\delta.$

§2 Setup

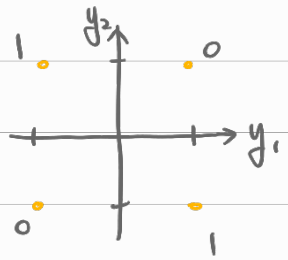
(1) $\mathcal{X}$ ··· input space ($\mathbb{R}^n$)

$\mathcal{Y}=\{\pm1\}$. $\mathcal{H}\subseteq[\mathcal{X}\to\mathcal{Y}]$

(2) $L\in[\mathcal{Y}\times\mathcal{Y}\to[0,1]]$. $L(y_1,y_2)=\mathbb{1}_{\{y_1\neq y_2\}}$.

(3) $\mathcal{G}\subseteq[\mathcal{X}\times\mathcal{Y}\to[0,1]]$

$= \{(x,y)\mapsto L(h(x),y)\mid h\in\mathcal{H}\}$. (Induced function class)



$\frac{1-y_1y_2}{2}$ (Bilinear extension).

§3 Rademacher Complexity

① Empirical Rademacher Complexity of $\mathcal{F}\subseteq[\mathcal{U}\to\mathcal{V}]$, $\mathcal{V}\subseteq\mathbb{R}$ with respect to $S=(u_1,u_2,\cdots,u_m)\in\mathcal{U}^m$ is

$$\hat{R}_S(\mathcal{F}) = \underset{\vec{\sigma}\in Unif(\{\pm1\}^m)}{\mathbb{E}}\left[\sup_{f\in\mathcal{F}} \frac{1}{m}\sum_{i=1}^{m}\sigma_i f(u_i)\right]$$

② **Rademacher Complexity** of $\mathcal{F}$ wrt $D\in Pr(\mathcal{U})$ and $m\in\mathbb{N}$ is

$$R_m(\mathcal{F}) = \mathbb{E}_{S\sim D^m}[\hat{R}_S(\mathcal{F})].$$

Exercise

(3) $\mathcal{F} = \{u\in\mathbb{R} \mapsto \text{if } u\in[a,b] \text{ then } +1 \text{ else } -1 \mid a,b\in\mathbb{R}\}$.

$S_0=(3)$, $S_1=(3,-2)$, $S_2=(3,-2,5)$, $S_4(3,3)$.

$\hat{R}_{S_i}(\mathcal{F})=?$

$\hat{R}_{S_0}(\mathcal{F})=1$

$\hat{R}_{S_1}(\mathcal{F})=1$

$\hat{R}_{S_2}(\mathcal{F})=\frac{7}{8}+\frac{1}{8}\times\frac{1}{3}=\frac{22}{24}=\frac{11}{12}$

$\hat{R}_{S_4}(\mathcal{F})=\frac{1}{2}$

-2 3 5

+1 -1 +1

-1 [+1] -1

② $D_0=\text{Dirac}_3$ $D_1=\text{Unif}[0,1]$

(1) $R_2(\mathcal{F})=\mathbb{E}_{S\sim D_0^2}[\hat{R}_S(\mathcal{F})]=\hat{R}_4(\mathcal{F})=\frac{1}{2}$.

(2) $R_2(\mathcal{F})=\mathbb{E}_{S\sim D_1^2}[\hat{R}_S(\mathcal{F})]=1$. ($\because \hat{R}_S(\mathcal{F})=1$ if S contains only two "distinct" points.)

Theorem 3.3 (Generalization Bound)

$\mathcal{F}\subseteq[\mathcal{U}\to[0,1]]$, $\delta>0$, $m\in\mathbb{N}$, $D\in Pr(\mathcal{U})$

① $\mathbb{P}_{S\sim D^m}\left[\forall f\in\mathcal{F}\ \mathbb{E}_{u\sim D}[f(u)]\le \mathbb{E}_{u\sim\hat{D}_S}[f(u)]+2R_m(\mathcal{F})+\sqrt{\frac{\log(1/\delta)}{2m}}\right]\ge 1-\delta$.

$\hat{D}_S \to \text{Unif}(S)$

② $\mathbb{P}_{S\sim D^m}\left[\forall f\in\mathcal{F}\ \mathbb{E}_{u\sim D}[f(u)]\le \mathbb{E}_{u\sim\hat{D}_S}[f(u)]+2\hat{R}_S(\mathcal{F})+3\sqrt{\frac{\log(2/\delta)}{2m}}\right]\ge 1-\delta$.

Theorem D.8 (McDiarmid's Inequality)

$X_1,\cdots,X_m$: Independent random variables. $x_i\in\mathcal{X}$.

$c_1,\cdots,c_m>0$.

$f:\mathcal{X}^m\to\mathbb{R}$ such that $|f(u_1,\cdots,u_m)-f(u_1,\cdots,u_i',\cdots,u_m)|\le c_i$ $\forall i$, $\forall u_1,\cdots,u_m,u_i'$.

Then, $\mathbb{P}[f(x_1,\dots,x_m) - \mathbb{E}[f(x_1,\dots,x_m)] \geq \varepsilon] \leq e^{-\frac{2\varepsilon^2}{\sum_{i=1}^m c_i^2}}$

$$\mathbb{P}[f(x_1,\dots,x_m) - \mathbb{E}[f(x_1,\dots,x_m)] \leq -\varepsilon] \leq e^{-\frac{2\varepsilon^2}{\sum_{i=1}^m c_i^2}}.$$

(Sketch of proof) $\Phi : \mathcal{U}^m \to [-1,1]$

$\Phi(S) = \sup_{f\in\mathcal{F}} \left(\mathbb{E}_{u\sim D}[f(u)] - \mathbb{E}_{u\sim \hat{D}_S}[f(u)] \right)$. It suffices to show:

① $\mathbb{P}_{S\sim D^m}\left[\Phi(S) \leq \mathbb{E}_{S\sim D^m}[\Phi(S)] + \sqrt{\frac{\log 1/\delta}{2m}} \right] \geq 1-\delta.$

② $\mathbb{P}_{S\sim D^m}[\Phi(S)] \leq 2R_m(\mathcal{F}).$

① By McDiarmid, with $c_i = \frac{1}{m}$, $f(S) = \frac{1}{m}\sum_{i=1}^m f(x_i)$,

then $\mathbb{P}\left[\Phi(S) \leq \mathbb{E}_{S\sim D^m}[\Phi(S)] + \sqrt{\frac{\log(1/\delta)}{2m}}\right] \geq 1 - \exp\left(-\frac{2\,\frac{\log 1/\delta}{2m}}{\sum_{i=1}^m \frac{1}{m^2}}\right)$

$$= 1-\delta.$$

$S = (u_1,\dots,u_m)$, $S' = (u_1,\dots,u_i',\dots,u_m)$.

$$\Phi(S) - \Phi(S') = \sup_{f\in\mathcal{F}} \quad - \sup_{f\in\mathcal{F}}$$

$$\leq \sup_{f\in\mathcal{F}} \mathbb{E}_{u\sim D}[f(u)] - \mathbb{E}_{u\sim\hat{D}_S}[f(u)] - \left(\mathbb{E}_{u\sim D}[f(u)] - \mathbb{E}_{u\sim \hat{D}_{S'}}[f(u)]\right)$$

$$= \sup_{f\in\mathcal{F}} \mathbb{E}_{u\sim\hat{D}_{S'}} f(u) - \mathbb{E}_{u\sim\hat{D}_S} f(u)$$

$$= \sup_f \frac{1}{m} f(u_i') - \frac{1}{m} f(u_i) \leq \frac{1}{m}.$$

By symmetry, $\Phi(S') - \Phi(S) \leq \frac{1}{m}$. Hence, $|\Phi(S) - \Phi(S')| \leq \frac{1}{m}$.

② $$\mathbb{E}_{S\sim D^m}[\Phi(S)] = \mathbb{E}_{S\sim D^m}\Big[\sup_{f\in\mathcal{F}}\big(\underbrace{\mathbb{E}_{u\sim D}[f(u)]}_{=\left(\mathbb{E}_{S'\sim D^m}\left[\mathbb{E}_{u'\sim \hat{D}_{S'}}[f(u')]\right]\right)} - \mathbb{E}_{u\sim \hat{D}_S}[f(u)]\big)\Big]$$

$$= \mathbb{E}_{S\sim D^m}\Big[\sup_{f\in\mathcal{F}} \mathbb{E}_{S'\sim D^m}\big[\mathbb{E}_{u'\sim \hat{D}_{S'}}[f(u')] - \mathbb{E}_{u\sim \hat{D}_S}[f(u)]\big]\Big]$$

$$\le \mathbb{E}_{S\sim D^m}\Big[\mathbb{E}_{S'\sim D^m} \sup_{f\in\mathcal{F}}\big[\mathbb{E}_{u'\sim \hat{D}_{S'}}[f(u')] - \mathbb{E}_{u\sim \hat{D}_S}[f(u)]\big]\Big]$$

$$= \mathbb{E}_{\substack{S,S' \\ \vec{\sigma}\in \mathrm{Unif}\{\pm1\}^m}}\Big[\sup_{f\in\mathcal{F}} \frac{1}{m}\sum_{i=1}^m \sigma_i(f(u_i) - f(u_i'))\Big]$$

$S = (u_1, \cdots, u_m)$, $S' = (u_1', \cdots, u_m')$

$$\vec{\sigma}(S) = \begin{cases} u_i & \text{if } \sigma(i) = 1 \\ u_i' & \text{otherwise} \end{cases} \qquad \vec{\sigma}(S') = \begin{cases} u_i' & \text{if } \sigma(i) = 1 \\ u_i & \text{otherwise.} \end{cases}$$

Then, $\mathbb{E}_{\substack{S\sim D^m, S'\sim D^m \\ \vec{\sigma}\in \mathrm{Unif}\{\pm1\}^m}}\Big[\qquad\qquad\Big]$

$$= \mathbb{E}_{\substack{\vec{\sigma}(S)\sim D^m, \vec{\sigma}(S')\sim D^m \\ \vec{\sigma}\in \mathrm{Unif}\{\pm1\}^m}}\Big[\sup \frac{1}{m}\sum_{i=1}^m \langle \vec{\sigma}, f(\vec{\sigma}(S)) - f(\vec{\sigma}(S'))\rangle\Big]$$

$$= \quad '' \quad \Big[\sup \frac{1}{m}\sum f(S) - f(S')\Big]$$

$$= \mathbb{E}_{S\sim D^m, S'\sim D^m}\Big[\sup \quad '' \quad\Big]$$

Hence, $$\mathbb{E}_{S\sim \Phi^m}[\Phi(S)] \le \mathbb{E}_{S,\vec{\sigma}}\Big[\sup_{f\in\mathcal{F}} \frac{1}{m}\sum_{i=1}^m \sigma_i f(u_i)\Big] + \mathbb{E}_{S,\vec{\sigma}}\Big[\sup_{f\in\mathcal{F}} \frac{1}{m}\sum_{i=1}^m \sigma_i f(u_i)\Big]$$

$$= R_m(\mathcal{F}) + R_m(\mathcal{F}) = 2R_m(\mathcal{F}).$$

§4. RC & Our Setup.

(1) $\mathcal{X}$, $\mathcal{Y} = \{\pm 1\}$, $\mathcal{H} \subseteq [\mathcal{X} \to \mathcal{Y}]$, $D \in Pr(\mathcal{X} \times \mathcal{Y})$.

$\mathcal{F} \subseteq [\mathcal{X} \times \mathcal{Y} \to [0,1]]$, $\mathcal{F} = \{(x,y) \mapsto \mathbb{1}_{h(x) \neq y} : h \in \mathcal{H}\}$.

(2) f_h

$$\mathbb{E}_{(x,y)\sim D}[f_h(x,y)] = \mathbb{E}_{(x,y)\sim D}[\mathbb{1}_{h(x)\neq y}] = R(h)$$

$$\mathbb{E}_{(x,y)\sim \hat{D}_S}[f_h(x,y)] = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{h(x_i)\neq y_i\}} = \hat{R}_S(h).$$

thm3.3 $\Rightarrow$ $\mathbb{P}_{S\sim D^m}\left[\overset{\forall h \in \mathcal{H}}{R(h) \le \hat{R}_S(h) + 2R_m(\mathcal{F}) + \sqrt{\frac{\log(1/\delta)}{2m}}}\right] \ge 1-\delta$

and so on.

Theorem3.4 $2\hat{R}_S(\mathcal{F}) = \hat{R}_{S_x}(\mathcal{H})$

$S = ((x_1,y_1),\cdots,(x_m,y_m))$, $S_x = (x_1,\cdots,x_m)$.

Here, thm3.3 $\mathbb{P}_{S\sim D^m}\left[\forall h \in \mathcal{H}\ \ R(h) \le \hat{R}_S(h) + R_m(\mathcal{H}) + \sqrt{\frac{\log(1/\delta)}{2m}}\right] \ge 1-\delta$

$\mathbb{P}_{S\sim D^m}\left[\ \text{"}\ \ R(h) \le \hat{R}_S(h) + \hat{R}_{S_x}(\mathcal{H}) + 3\sqrt{\frac{\log(2/\delta)}{2m}}\right] \ge 1-\delta.$

(pf of theorem3.4)

$$\hat{R}_S(\mathcal{F}) = \mathbb{E}_{\sigma}\left[\sup_{f\in\mathcal{F}} \frac{1}{m}\sum_{i=1}^{m} \sigma_i f(x_i,y_i)\right]$$

$$= \mathbb{E}_{\sigma}\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum \sigma_i \mathbb{1}_{h(x_i)\neq y_i}\right]$$

$$= \mathbb{E}_{\sigma}\left[\sup_{h} \frac{1}{m}\sum \frac{\sigma_i}{2}(1 - h(x_i)y_i)\right]$$

$$= \mathbb{E}_{\sigma}\left[\sup_{h} \frac{1}{m}\sum \frac{\sigma_i}{2} - \frac{1}{m}\sum \frac{\sigma_i}{2} h(x_i)y_i\right]$$

$$= \mathbb{E}_{\sigma}\left[\cancelto{0}{\frac{1}{m}\sum\frac{\sigma_i}{2}} + \sup_{h} \frac{1}{m}\sum \frac{1}{2}(-\sigma_i y_i)h(x_i)\right]$$

$$= \frac{1}{2}\mathbb{E}_{\sigma}\left[\sup_{h} \frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)\right]$$

$$= \frac{1}{2}\hat{R}_{S_x}(\mathcal{H}).$$

Sep / 19 / 2023

Notes

- Growth Function : $\Pi_{\mathcal{H}} \in [\mathbb{N} \to \mathbb{N}]$

$$R_m(\mathcal{H}) \leq \sqrt{\frac{2\log \Pi_{\mathcal{H}}(m)}{m}}$$

- VC-dimension : $VCDim(\mathcal{H}) = d$.

$$\Pi_{\mathcal{H}}(m) \leq \left(\frac{em}{d}\right)^d \text{ for all } m \geq d$$

§ Growth Function

Definition (Growth Function)

$\Pi_{\mathcal{H}}$ of a hypo set $\mathcal{H} \subseteq [\mathcal{X} \to \underset{\{\pm 1\}}{\underset{\shortparallel}{\mathcal{Y}}}]$

is $\Pi_{\mathcal{H}}(m) = \max_{(x_1,\cdots,x_m) \in \mathcal{X}^m} |\{(h(x_1),\cdots,h(x_m)) \mid h \in \mathcal{H}\}|$.

- $\Pi_{\mathcal{H}}(m) \leq 2^m$
- $\Pi_{\mathcal{H}}(m) \leq \Pi_{\mathcal{H}'}(m)$ for $\mathcal{H} \subseteq \mathcal{H}'$.

Theorem 3.7 (Massart's Lemma)

Let $A \subseteq \mathbb{R}^m$ be a finite nonempty set s.t. $r = \max_{y \in A} \|y\|_2 = \sqrt{\sum_{i=1}^m y_i^2}$.

Then, $\mathbb{E}_{\vec{\sigma} \sim Unif\{\pm1\}^m}\left[\max_{\underbrace{y \in A}_{h \in \mathcal{H}}} \frac{1}{m}\sum_{i=1}^m \sigma_i \underbrace{y_i}_{h(x_i)}\right] \leq \frac{1}{m}\left(r\sqrt{2\log|A|}\right)$.

Corollary 3.8 $R_m(\mathcal{H}) \leq \sqrt{\frac{2\log \Pi_{\mathcal{H}}(m)}{m}}$

(pf) $R_m(\mathcal{H}) = \mathbb{E}_{S \sim D^m}[\hat{R}_{S_x}(\mathcal{H})]$, Let $y_h = (h(x_1),\cdots,h(x_m)) \Rightarrow \|y_h\|_2 = \sqrt{m}\ \forall h$

$\Rightarrow r = \sqrt{m}$

$|\underbrace{\{y_h \mid h \in \mathcal{H}\}}_{A}| \leq \Pi_m(\mathcal{H})$

$\hat{R}_{S_x}(\mathcal{H}) = \mathbb{E}_{\vec{\sigma}}\left[\sup_{h \in \mathcal{H}} \frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i)\right]$

$$= \mathbb{E}_{\vec{\sigma}}\left[\sup_{y \in A} \frac{1}{m}\sum_{i=1}^{m} \sigma_i y_i\right] \le \frac{1}{m}\left(r\sqrt{2\log|A|}\right)$$

$$\le \frac{1}{m}\left(r\sqrt{2\log \Pi_m(\mathcal{H})}\right) = \sqrt{\frac{2\log \Pi_m(\mathcal{H})}{m}}.$$

Hence, $\hat{R}_m(\mathcal{H}) \le \sqrt{\frac{2\log \Pi_m(\mathcal{H})}{m}}$.

§2 VC Dimension

Definition (Shattered Domain)

Let $\mathcal{X}$, $\mathcal{Y} = \{\pm 1\}$, $\mathcal{H} \subseteq [\mathcal{X} \to \mathcal{Y}]$.

A subset $X_0 \in \mathcal{X}$ is shattered if for any $X_0' \subseteq X_0$, $\exists h \in \mathcal{H}$ s.t.
$X_0' = \{x \in X_0 \mid h(x) = 1\}$.

Definition (VC Dimension)

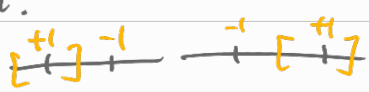
VC Dimension of $\mathcal{H}$ is the size of maximal shattered set of inputs.

$$\mathrm{VCDim}(\mathcal{H}) = \sup_{\substack{X_0 \subseteq_{fin} \mathcal{X} \\ X_0 \text{ shattered by } \mathcal{H}.}} |X_0|$$

$$= \sup\{m \mid \Pi_{\mathcal{H}}(m) = 2^m\}.$$

Example 1.

$\mathcal{X} = \mathbb{R}$, $\mathcal{H} = \{x \mapsto -1 + 2\cdot \mathbb{1}_{[l,r]} \mid l, r \in \mathbb{R}\}$.

Any $X_0 \subseteq \mathcal{X}$ of $|X_0| = 2$ is shattered by $\mathcal{H}$. 

Any $X_0 \subseteq \mathcal{X}$ of $|X_0| = 3$ is not shattered by $\mathcal{H}$. 

$\Rightarrow$ No possibility of choosing $h \in \mathcal{H}$.

Thus, $\mathrm{VCDim}(\mathcal{H}) = 2$

2. $\mathcal{X} = \mathbb{R}^n$, $\mathcal{H} = \{x \longmapsto \mathrm{sgn}(\langle w, x\rangle + b) \mid w \in \mathbb{R}^n, b \in \mathbb{R}\}$. $\mathrm{sgn}(u) = \begin{cases} +1 & \text{if } u \ge 0 \\ -1 & \text{otherwise} \end{cases}$.

$\mathrm{VCDim}(\mathcal{H}) \ge n+1$. Take $n+1$ points of $\mathbb{R}^n = \{0, e_1, \cdots, e_n\} =: X_0$

$y_0, y_1, \cdots, y_n$

$n = 2 \Rightarrow$  $\Rightarrow \mathrm{VCDim}(\mathcal{H}) = 3$ for $n = 2$.

If for $X_0' \subseteq X_0$, Let $\sigma_i = \begin{cases} +1, & \text{if } 0, e_i \in X_0' \\ -1 & \text{if } 0, e_i \notin X_0'. \end{cases}$

Then, $y_0 = \text{sgn}(b)$.

$y_i = \text{sgn}(w_i + b)$.

$b = y_0/2$, $w_i = y_i$.

Theorem (Radon's Theorem)

If $X \subseteq_{fin} \mathbb{R}^d$ with $|X| = d+2$. Then, $\exists\ X_0, X_1 \subseteq X$ s.t.

$\underbrace{X_0 \cap X_1 = \emptyset;\ X_0 \cup X_1 = X,\ X_0 \neq \emptyset,\ X_1 \neq \emptyset}_{X_0, X_1 \text{ is partition}}$, $\text{conv}(X_0) \cap \text{conv}(X_1) \neq \emptyset$

(pf) $X_1 = \{x_1, \dots, x_{d+2}\}$, $x_i \in \mathbb{R}^{d+2}$

Take $\alpha_1, \dots, \alpha_{d+2} \in \mathbb{R}$ s.t. $\sum_{i=1}^{d+2} \alpha_i = 0$, $\sum_{i=1}^{d+2} \alpha_i x_i = 0$. $\longrightarrow$ $d+1$ constraint.

$\exists$ solution that is nontrivial. ($\exists j$ s.t. $\alpha_j \neq 0$ for some j)

$J_1 = \{i \in [d+2] \mid \alpha_i > 0\}$

$J_2 = \{i \in [d+2] \mid \alpha_i \leq 0\}$.

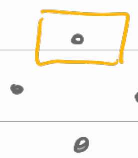
Then, J_1, J_2 are nonempty, as $\sum \alpha_i = 0$.

Let $X_1 = \{x_i \mid i \in J_1\}$, $X_2 = \{x_i \mid i \in J_2\}$. Then, $\alpha = \sum_{i \in J_1} \alpha_i > 0$.

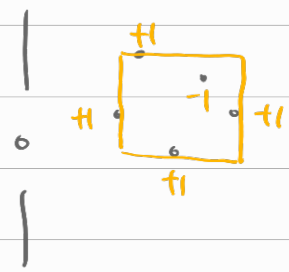
$$\sum_{i \in J_1} \alpha_i x_i + \sum_{i \in J_2} \alpha_i x_i = 0 \Rightarrow \underset{\in\, \text{conv}(X_1)}{\sum_{i \in J_1} \frac{\alpha_i}{\alpha} x_i} = \underset{\in\, \text{conv}(X_2)}{\sum_{i \in J_2} -\frac{\alpha_i}{\alpha} x_i}.$$

Example 1. (Lower Bound of VCDim($\mathcal{H}$))

$\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \left\{ x \mapsto \begin{cases} +1 & \text{if } x \in [l, r] \times [b, t] \\ -1 & \text{otherwise} \end{cases} \;\middle|\; l, r, b, t \in \mathbb{R} \right\}$

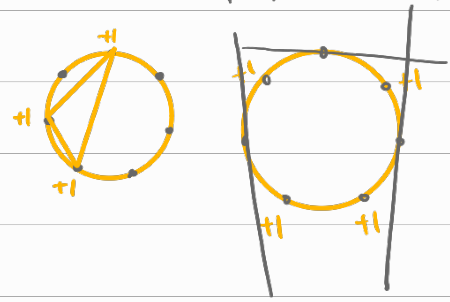


$\Rightarrow \text{VCDim}(\mathcal{H}) \geq 4$



2. $\mathcal{X}=\mathbb{R}^2$. $\mathcal{H}$ is the set of d-dimensional convex polygon.

$d=3$



→ shrink Δ a little bit.

3. $\mathcal{X}=\mathbb{R}$, $\mathcal{H}=\{x\mapsto \mathrm{sgn}(\sin(wx)) \mid w\in\mathbb{R}\}$, $d=\infty$

§3 Generalization Bound via VC Dimension.

(1) Goal $\quad \Pi_{\mathcal{H}}(m)\le\left(\frac{em}{d}\right)^d$, where $d=\mathrm{VCDim}(\mathcal{H})$

Theorem 3.17 (Sauer's Lemma)

$$\Pi_{\mathcal{H}}(m)\le\sum_{i=0}^{d}\binom{m}{i}$$
$$\rotatebox{90}{$\le$}$$
$$2^m=\sum_{i=0}^{d}\binom{m}{i}$$

(pf) By induction on m+d.

	m=1	m=2	m=3	...
d=0				
d=1				
d=2				
⋮				

i) $d=0 \Rightarrow |\mathcal{H}|=1$. Thus, $\Pi_{\mathcal{H}}(m)=1\le\sum_{i=0}^{d}\binom{m}{i}=1$.

$m=1, d\neq 0 \Rightarrow \Pi_{\mathcal{H}}(m)\le 2^m=2=\binom{m}{0}+\binom{m}{1}\le\sum_{i=0}^{d}\binom{m}{i}$

ii) Inductive case $m\ge 2, d\ge 1$

$$\Pi_{\mathcal{H}}(m)=\max_{S=(x_1,\dots,x_m)}\left|\{(h(x_1),\dots,h(x_m)) \mid h\in\mathcal{H}\}\right|$$

$S\subseteq_{fin}\mathcal{X}$. $\mathcal{H}|_S=\{h|_S\in[S\to\mathcal{Y}] \mid h\in\mathcal{H}\}$

$\exists$ subset $S_0\subseteq_{fin}\mathcal{X}$ s.t. $|S_0|=m$, $\Pi_{\mathcal{H}}(m)=|\mathcal{H}|_{S_0}|$.

Let $S_0 = \{x_1, \cdots, x_m\}$, $S_1 := \{x_1, \cdots, x_{m-1}\}$.

$\mathcal{H}|_{S_1}$, $\mathcal{H}' := \{h \in [S_1 \to \mathcal{Y}] \mid \exists h_1, h_2 \in \mathcal{H}|_{S_0}$ s.t. $h_1(x_m) = +1, h_2(x_m) = -1,$

$h = h_1|_{S_1} = h_2|_{S_1}\}$

$$\text{Then, } \Pi_{\mathcal{H}}(m) = |\mathcal{H}|_{S_0}| = |\overset{\text{VCDim} \le d}{\mathcal{H}|_{S_1}}| + |\overset{\text{VCDim} \le d-1}{\mathcal{H}'}|.$$

$$= \Pi_{\mathcal{H}|_{S_1}}(m-1) + \Pi_{\mathcal{H}'}(m-1).$$

$$\le \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d-1} \binom{m-1}{i} \quad \text{(By induction hypothesis)}$$

$$\le \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d} \binom{m-1}{i-1}$$

$$= \sum_{i=0}^{d} \binom{m}{i}$$

It remains to check $\text{VCDim}(\mathcal{H}|_{S_1}) \le d$, $\text{VCDim}(\mathcal{H}') \le d-1$.

If a subset $X_0 \subseteq S_1$ is shattered by $\mathcal{H}'$,

$X_0 \cup \{x_m\} \subseteq S_0$ is shattered by $\mathcal{H}|_{S_0}$

Thus, $\text{VCDim}(\mathcal{H}') + 1 \le \text{VCDim}(\mathcal{H}|_{S_0}) \le \text{VCDim}(\mathcal{H}) = d$.

Sep / 26 / 2023

§ **Recall**

$$\mathbb{P}_{S\sim D^m}\left[\ \forall h\in\mathcal{H}\quad R(h)\le \hat{R}_S(h)+R_m(\mathcal{H})+\sqrt{\frac{\log(1/\delta)}{2m}}\ \right]\ge 1-\delta.$$

$$R_m(\mathcal{H})\le\sqrt{\frac{2\log\Pi_{\mathcal{H}}(m)}{m}}\le\sqrt{\frac{2\log(em/d)}{m/d}},\quad d=\mathrm{VCDim}(d).$$

$$\Pi_{\mathcal{H}}(m)\le\left(\frac{em}{d}\right)^d.$$

Theorem 3.17 (Sauer's Lemma)

$$\Pi_{\mathcal{H}}(m)\le\sum_{i=0}^{d}\binom{m}{i}.$$

Corollary 3.18 $\Pi_{\mathcal{H}}(m)\le\left(\frac{em}{d}\right)^d.\quad \forall m\ge d$

(pf of cor.)

$$\begin{aligned}\Pi_{\mathcal{H}}(m)&\le\sum_{i=0}^{d}\binom{m}{i}\quad(\text{by Sauer's lemma})\\&\le\sum_{i=0}^{d}\binom{m}{i}\left(\frac{m}{d}\right)^{d-i},\quad m\ge d.\\&\le\sum_{i=0}^{m}\binom{m}{i}\left(\frac{m}{d}\right)^{d-i}\\&=\left(\frac{m}{d}\right)^d\sum_{i=0}^{m}\binom{m}{i}\left(\frac{d}{m}\right)^i\\&=\left(\frac{m}{d}\right)^d\left(1+\frac{d}{m}\right)^m\\&\le\left(\frac{m}{d}\right)^d e^d=\left(\frac{em}{d}\right)^d.\end{aligned}$$