1. HW1 is up

(problem 4.1. last term in second equation

$$\frac{6 \ln \frac{2|\mathcal{H}|}{\delta}}{m} \to \frac{12 \ln \frac{2|\mathcal{H}|}{\delta}}{m} \quad)$$

Sauer's Lemma

$\mathcal{H}$: nonempty $\quad S = (x_1 \ldots x_n)$ unlabeled examples

$$|\Pi_{\mathcal{H}}(S)| \leq |\{ T \subseteq S : \mathcal{H} \text{ shatters } T \}|$$

(If $VC(\mathcal{H}) = d$. $RHS \leq \sum_{i=0}^{d} \binom{n}{i} =: \binom{n}{\leq d}$)

[The converse is true up to log factors :
$\mathcal{H}$. $S(\mathcal{H}, n) \leq \sum_{i=0}^{d} \binom{n}{i}$ for all n $\Rightarrow VC(\mathcal{H}) \leq O(d \ln d)$]

pf : Induction on n.

Base case : $n = 1$

$\mathcal{H}$	x_1
h	$+1$

① $\mathcal{H}$ agrees on x_1

$LHS = 1 \quad RHS = 1$. only ϕ is shattered by $\mathcal{H}$.

② $LHS = 2$

$\mathcal{H}$	x_1
h_1	$+1$
h_2	-1

$RHS = 2$.
both ϕ and $\{x_1\}$ shattered by $\mathcal{H}$.

①② $\Rightarrow$ LHS $\leq$ RHS

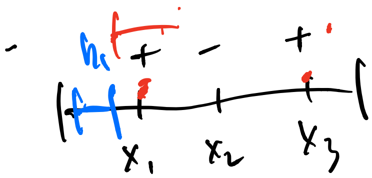
Inductive case: $n \geq 2$

inductive hypothesis: for all S' of size $n-1$, $|\Pi_{\mathcal{H}}(S')| \leq |\{T \subseteq S' : \mathcal{H} \text{ shatters } T\}|$

upper bound $\Pi_{\mathcal{H}}(S)$ $\quad |S| = n$.

$\mathcal{H}_S$: for every labeling $(l_1 \cdots l_n)$ in $\Pi_{\mathcal{H}}(S)$, select one representative in $\mathcal{H}$, add it to $\mathcal{H}_S$.

$$|\mathcal{H}_S| = |\Pi_{\mathcal{H}}(S)|$$



Interval class

$\{2\mathbb{I}(x \in [a,b]) - 1 : a, b \in \mathbb{R}\}$

$\mathcal{H}_S$	x_1	x_2	x_3
h_1	−	−	−
h_2	+	−	−
h_3	−	+	−
h_4	−	−	+
h_5	+	+	−
h_6	−	+	+
h_7	+	+	+

3 4

Decomposition of $\mathcal{H}_S$: $\quad S' = \{x_1 \cdots x_{n-1}\}$ $\quad S = S' \cup \{x_n\}$

$= S \setminus \{x_n\}$

For every labeling on S': $(l_1 \cdots l_{n-1})$

① if both $(l_1 \cdots l_{n-1}, +1)$ and $(l_1 \cdots l_{n-1}, -1)$ are achievable by $\mathcal{H}$, send one classifier to $\mathcal{H}_1$, send the other to $\mathcal{H}_2$

② if only one of $(l_1 \cdots l_{n-1}, +1)$ and $(l_1 \cdots l_{n-1}, -1)$ is achievable by $\mathcal{H}_S$.

$S = \{x_n\}$

$\mathcal{H}_1$

	x_1	x_2	x_3
h_1	−	−	−
h_2	+	−	−
h_3	−	+	−
h_5	+	+	−

$\mathcal{H}_2$

	x_1	x_2	x_3
h_4	−	−	+
h_6	−	+	+
h_7	+	+	+

$|\mathcal{H}_1| \geq |\mathcal{H}_2|$

Observations: ① $|\mathcal{H}_1| = |\Pi_{\mathcal{H}_1}(S')|$

$|\mathcal{H}_2| = |\Pi_{\mathcal{H}_2}(S')|$

(all classifiers in $\mathcal{H}_1$ / $\mathcal{H}_2$ generate unique labelings in S')

② if $T \subseteq S'$ and $\mathcal{H}_1$ shatters T, then $\mathcal{H}_S$ shatters T.

e.g. $\mathcal{H}_1$ shatters $\{x_1, x_2\}$ $\Rightarrow$ $\mathcal{H}_S$ shatters $\{x_1, x_2\}$

③ if $T \subseteq S'$. and $\mathcal{H}_2$ shatters T, then

$\mathcal{H}_S$ shatter $T \cup \{x_n\}$

$\mathcal{H}_2$ shatters $\{x_2\}$ $\Rightarrow$ $\mathcal{H}_S$ shatters $\{x_2, x_3\}$

If $\mathcal{H}_2$ achieves $b_1 \dots b_{|T|}$ on T

then $\mathcal{H}$ will achieve both

$$(b_1 \dots b_{|T|}, +1)$$

and $(b_1 \dots b_{|T|}, -1)$

on $T \cup \{x_n\}$.

Concluding the proof:

$$|\mathcal{H}_S| = |\mathcal{H}_1| + |\mathcal{H}_2|$$

size $n-1$

$$\overset{①}{=} |\Pi_{\mathcal{H}_1}(S')| + |\Pi_{\mathcal{H}_2}(S')|$$

Inductive hyp.

$$\leq |\{T \subseteq S' : T \text{ shattered by } \mathcal{H}_1\}|$$

$$+ |\{T \subseteq S' : T \text{ shattered by } \mathcal{H}_2\}|$$

② ③

$$\le (*1)\ |\{T \subseteq S' : T \text{ shattered by } H_S\}|$$

$$+\ |\{T \subseteq S' : T \cup \{x_n\} \text{ shattered by } H_S\}|$$

$\{x_2\}$

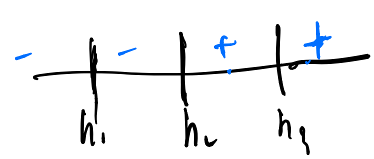
$\{x_2, x_3\}$

$$\|$$

$$(*2)\ |\{T \subseteq S : x_n \in T,\ T \text{ shattered by } H_S\}|$$

$$= |\{T \subseteq S : T \text{ shattered by } H_S\}|$$

Application of Sauer's lemma for bounding VC dimensions of hypothesis classes. H = threshold



maj(h1, h2, h3)

Ex : H : $VC(H) = d$.

k : odd number

$H_{f,k} = \{ f$

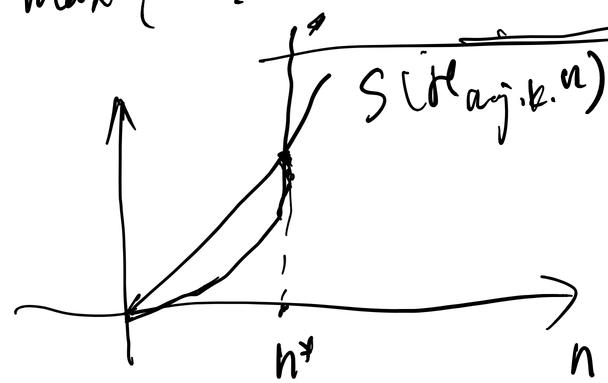
$$H_{maj,k} = \{ maj(h_1(x), \dots h_k(x)) : h_1 \dots h_k \in H \}$$

$$\mathrm{maj}(y_1 \cdots y_k) = \begin{cases} +1 : & |\{i : y_i = +1\}| > \frac{k}{2} \\ -1 & \end{cases}$$

can we bound $VC(\underset{\mathcal{H}_{f,k}}{\mathcal{H}_{maj,k}}) \leq \tilde{O}(k \cdot d)$?

Claim $S(\mathcal{H}_{maj,k}, n) \leq n^{k(d+1)}$

$$VC(\mathcal{H}_{maj,k}) = \max\{n : 2^n = S(\mathcal{H}_{maj,k}, n)\}$$



$\mathcal{H}$	x_1	x_2	$\cdots$	x_n		
h_1	$+1$	-1		$+1$	$\leq n^{d+1}$ (Sauer)	$k=1$
$\vdots$					$(n^{d+1})^k$	$k=3$
h_k	-1	$+1$		$+1$		

$\Downarrow$

$\underset{f}{\mathrm{maj}(h_1 \cdots h_k)}$ [. − − .] $\leq (n^{d+1})^k$ labelings.

n: $2^n \le n^{(d+1)k}$. solve this ineq.

$\Rightarrow n \le (d+1)\cdot k \cdot \log_2 n$

$b=0$

a

$\Rightarrow n \le \underbrace{2(d+1)\cdot k}_{} \ln n \Rightarrow n \le 4(d+1)\cdot k \cdot \ln(4(d+1)k) = \tilde{O}(d\cdot k)$

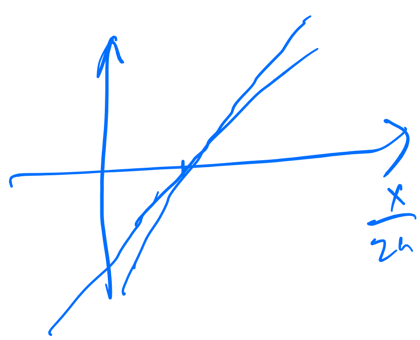
Useful Lemma $a \ge \frac{1}{2}$, $b \ge 0$.

$x \le a \ln x + b \Rightarrow x \le 2a\ln(2a) + 2b$.

This proves that $VC(\mathcal{H}_{maj,k}) = \tilde{O}(d\cdot k)$.

pf of Lemma:

$\ln \frac{x}{2a} \le \frac{x}{2a} - 1$



$\Rightarrow \ln x \le \frac{x}{2a} + \ln(2a)$

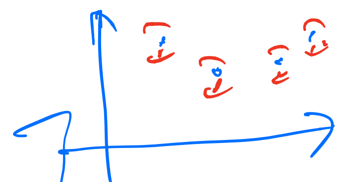
$x \le a \ln x + b$

$$\le \frac{x}{2} + a\ln(2a) + b$$

$$\frac{x}{2} \le a\ln(2a) + b$$

$$x \le 2a\ln(2a) + 2b$$

Theorem: suppose $\mathcal{H}$, $VC(\mathcal{H}) = d$. Then, given a set of n i.i.d. training examples $(x_1, y_1) \dots (x_n, y_n)$ from D, with prob. $1-\delta$,



$$\sup_{h\in\mathcal{H}} \left| err(h,S) - err(h,D) \right| \le c_1 \cdot \sqrt{\frac{d\ln\frac{n}{d} + \ln\frac{1}{\delta}}{n}} \quad (*)$$

for some constant c_1.

$\Downarrow$?

Consequently, ERM on $\mathcal{H}$ has an agnostic PAC sample complexity of (Exercise)

$$f(\varepsilon, \delta) = O\left(\frac{1}{\varepsilon^2}\left(d\ln\frac{1}{\varepsilon} + \ln\frac{1}{\delta}\right)\right).$$

population minimizer

$$(\ast) \Rightarrow \underbrace{\text{err}(\hat{h}, D) - \min_{h' \in H} \text{err}(h', D)}_{\text{excess error}}$$

$$\leq \underbrace{\tilde{O}\left(\sqrt{\frac{d}{n}}\right)}_{} \leq \varepsilon$$

when $n \geq \tilde{O}\left(\frac{d}{\varepsilon^2}\right)$, the RHS $\leq \varepsilon$.

ERM: $\hat{h} = \arg\min_{h \in H} \text{err}(h, S)$