Chapter 12 Minimax lower bounds

By fundamental theorem of statistical learning

suppose

$\{X, Y=\{0,1\}, \mathcal{F}, P\}$ $\quad V = VC(\mathcal{F}) < +\infty$

ERM: $\quad \hat{f}_n = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} 1_{\{f(X_i) \neq Y_i\}}$

wp $\geq 1-\delta$:

$$L(\hat{f}_n) \leq L^*(\mathcal{F}) + C\left(\sqrt{\frac{V}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right)$$

$(h=0)$

Also $\quad E[L(\hat{f}_n) - L^*(\mathcal{F})] \leq C\sqrt{\frac{V}{n}} \quad (C')$

expected excess risk

So $\quad \sup_P E[L(\hat{f}_n) - L^*(\mathcal{F})] \leq C\sqrt{\frac{V}{n}}$

min max expected excess risk:

$$\min_{\hat{f} = A(Z^n)} \sup_P E[L(\hat{f}_n) - L^*(\mathcal{F})]$$

$$L^*(\mathcal{F}) = \min_{f \in \mathcal{F}} L(f)$$

$$L(f) = P\{Y \neq f(X)\}$$

For P fixed:

$$\eta(x) = P\{Y=1 \mid X=x\}$$

$$f^*(x) = \begin{cases} 1 & \text{if } \eta(x) \geq 1/2 \\ 0 & \text{if } \eta(x) < 1/2 \end{cases}$$

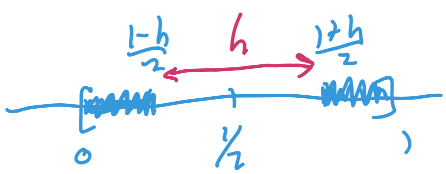
Deterministic label case: $\eta(x) \in \{0,1\}$ all x.

Can be shown $\sup_P E[L(\hat{f}_n) - L^*] \leq c\frac{V}{n}$ $(h=1)$

in deterministic label case

gap between upper bound $c\frac{V}{n}$ in special case and $\sqrt{\frac{V}{n}}$ if η is not constrained

Let $\mathcal{P}(h,\mathcal{F}) = \{P \in \mathcal{P}(\mathcal{F}) : \underline{|2\eta(X)-1| \geq h} \text{ a.s.}\}$



$\eta(x) \in [\frac{1-h}{2}, \frac{1+h}{2}]$

Let $R_n(h,\mathcal{F}) = \inf_{\hat{f}_n} \sup_{P \in \mathcal{P}(h,\mathcal{F})} E\left[L(\hat{f}_n) - L(f^*)\right]$

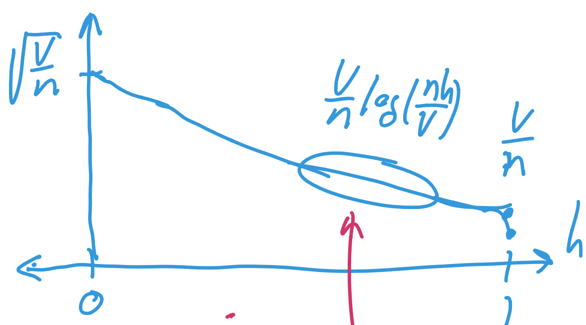
[$\mathcal{P}(\mathcal{F})$ = set of probability dist$^{\underline{n}}$ on X
such that f^* is in $\mathcal{F}$,
($f^*(x) = \mathbb{1}_{\{\eta(x) \geq 1/2\}}$, $\eta(x) = E(Y \mid X = x)$)]

<u>Theorem 12.1</u> (we'll prove) $\quad$ ($c = 1/32$ works)

$$R_n(h,\mathcal{F}) \geq c \min\left(\sqrt{\frac{V}{n}}, \frac{V}{nh}\right)$$

(Upper bound on <u>ERM</u>:

$$\sup_{P \in \mathcal{P}(h,\mathcal{F})} E\left[L(\hat{f}_n) - L^*\right] \leq \begin{cases} c\sqrt{\frac{V}{n}} & \text{if } h \leq \sqrt{V/n} \\ c\frac{V}{nh}\left(1 + \log\left(\frac{nh^2}{V}\right)\right) & \text{if } h > \sqrt{V/n} \end{cases}$$



we have matching lower bound if the sets of classifiers is sufficiently rich (beyond VC dimension)

Why VC dimension may not be enough.

Example A (Fix $d \geq 1$) (Not (N,D) rich for any $N \geq d+1$)

$X = \{1, \cdots, d\}$, $\mathcal{F}$ = set of all binary valued functions on X.

VC dim. is $V = d$

0 0 1 1 0 1 0 1

$\longleftarrow d \longrightarrow$

Example B

$X = \mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$

$\mathcal{F}$ = set of all binary valued functions such that $\sum_{x \in \mathbb{Z}} f(x) \leq d$

f: 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0

$\cdots$ -2 -1 0 1 2 3 $\cdots$

$V = d$ (Is (N, D) rich for and D with $0 \leq D \leq d$.

Definition $(X, Y = \{0,1\}, \mathcal{F}, P)$

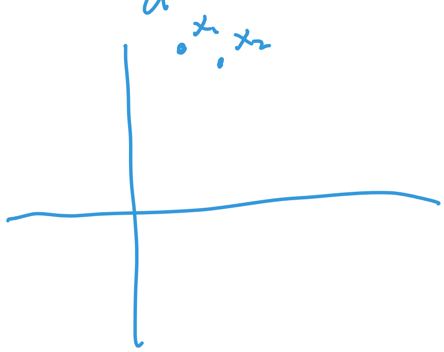
We say $\mathcal{F}$ is (N, D) rich for some $N \geq 1, D \geq 1$ if there exist $x_1, \cdots, x_N \in X$ so that for any length N binary sequence with D ones, b there exists $f \in \mathcal{F}$ so $(f(x_1), \cdots, f(x_N)) = (b_1, \cdots, b_N)$

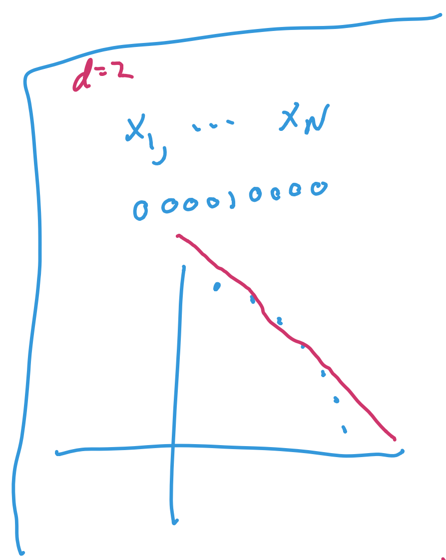
If $V = $ VC dimension$(\mathcal{F})$ then
$\mathcal{F}$ is (V, D)-rich for any D with $0 \le D \le V$.

Example $\mathcal{F}$ = set of halfspaces with boundary not necessarily through the origin in $\mathbb{R}^d$.

$V(\mathcal{F}) = d$ $(N, 1)$ rich

$d = 2$,





$VC(\mathcal{F}) = d$

($\mathcal{F}$ is $(N, \lfloor d/2 \rfloor)$-rich for all $N \ge d+1$)

Theorem 12.2 Given some $D \ge 1$, suppose $\mathcal{F}$ is (N, D) rich for all $N \ge 4D$. Then

$$R_n(h, \mathcal{F}) \ge c(1-h)\frac{D}{nh}\left[1 + \log \frac{nh^2}{D}\right]$$