Chapter 8 Binary Classification

$(X, \mathcal{C}, P)$ $\quad P \in \mathcal{P}$ $\quad \underline{P}$ is a prob dist^n on $X \times \{0,1\}$ – $\quad Z_i = (X_i, Y_i)$

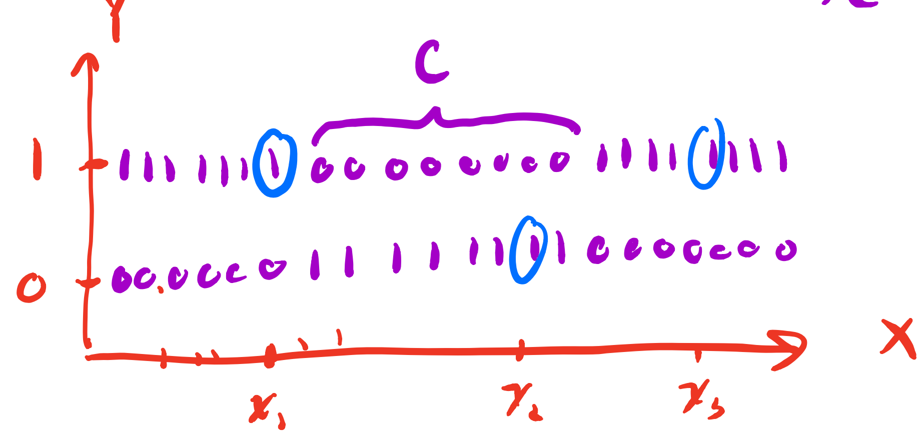
To use the most abstract framework $(Z, \mathcal{F}, \underline{P})$ we let $f_c(\underset{x,y}{z}) = 1_{\{y \neq 1_c(x)\}}$

$$\mathcal{F} = \mathcal{F}_{\mathcal{C}} = \{f_c : c \in \mathcal{C}\}$$

$$E[\Delta(Z_n)] \le 2E[R_n(\mathcal{F}(Z^n))] \le 4\sqrt{\frac{\overset{V(\mathcal{F}_{\mathcal{C}})}{V(\mathcal{F})}\log(n+1)}{n}}$$

Lemma 8.1 $\quad V(\mathcal{C}) = V(\mathcal{F}_C)$ $\quad z = (x, y)$

Proof $\quad z_1, z_2, z_3$ $\quad f_c$



Thm 8.1 $\quad m$ for Dudley class.

$$L_P(\hat{C}_n) \le L_P^*(\mathcal{C}) + \varepsilon\sqrt{\frac{VC \log(n+1)}{n}} + \sqrt{\frac{2\log(1/\delta)}{n}}$$

$\hat{C}_n$: ERM

$\langle w, z\rangle + b \ge 0$

e.g. $\mathcal{C} = \mathrm{pos}(g+h) = \{\{z : g(z)+h(z) \ge 0\}\}$

$$g(z) = \sum_{j=1}^{m} c_j \psi_j(z)$$

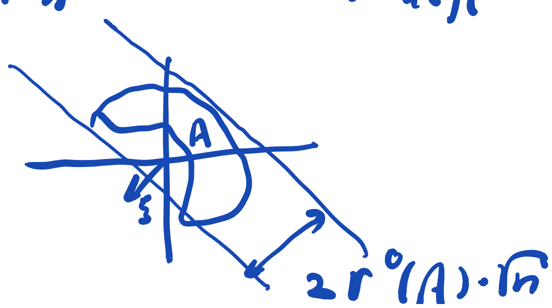
Dudley class $\quad V(\mathcal{C}) = m$

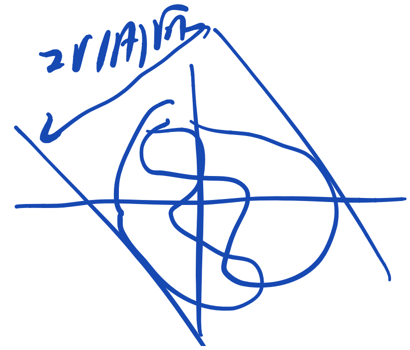
Revisit Rademacher averages

Let $A \subset \mathbb{R}^n$ bounded

$$R_n^0(A) = \frac{1}{n} E_{\varepsilon^n} \sup_{a\in A} \sum_{i=1}^{n} \varepsilon_i a_i$$

$$R_n(A) = \frac{1}{n} E_{\varepsilon^n} \sup_{a\in A} \left|\sum_{i=1}^{n} \varepsilon_i a_i\right| = R_0(A \cup -A)$$





Contraction principle for R. A.

Suppose $\varphi_i : \mathbb{R} \to \mathbb{R}$ for $i \in [n]$

let $\varphi \circ a = (\varphi_1(a_1), \varphi_2(a_2), \dots, \varphi_n(a_n))$

Suppose φ_i is M-Lipschitz continuous

i.e $|\varphi_i(a_i) - \varphi_i(a_i')| \leq M |a_i - a_i'|$

Let $\varphi \circ A = \{ \varphi \circ a : a \in A \}$

Then $\underline{R_n^{\circ}(\varphi \circ A) \leq M R_n^{\circ}(A)}$

If $\underline{\varphi_i(0) = 0}$ for all i (in addition to above)

then $R_n(\varphi \circ A) \leq 2M R_n(A)$

<u>Proof</u> (6.27) Let $M=1$ wlog

Let $\varphi_j(a_j) = a_j$ $j \neq 1$ wlog

$|\varphi_1(a_1) - \varphi_1(a_1')| \leq |a_1 - a_1'|$

$\varepsilon_1 = 1$ wp $1/2$

$ -1$ " $1/2$

$$-R_n^\circ(A) = \frac{1}{2n} E\left[\sup_a \left(a_1 + \sum_{i=2}^n \varepsilon_i a_i\right) + \sup_{a'}\left(-a_1' + \sum_{j=2}^n \varepsilon_i a_i'\right)\right]$$

$$= \frac{1}{2n} E\left[\sup_a \sup_{a'} \; a_1 - a_1' + \sum_{i=2}^n \varepsilon_i a_i + \sum_{j=2}^n \varepsilon_i a_i'\right]$$

$$= \frac{1}{2n} E\left[\sup_a \sup_{a'} |a_1 - a_1'| + \sum_{i=2}^n \varepsilon_i a_i + \sum_{j=2}^n \varepsilon_i a_i'\right]$$

$$\lor$$

$$-R_n^\circ(\varphi_i \circ A) = \frac{1}{2n} E\left[\sup_a \sup_{a'} |\varphi_1(a_1) - \varphi_1(a_1')| + \sum_{i=2}^n \varepsilon_i a_i + \sum_{j=2}^n \varepsilon_i a_i'\right]$$

$$\therefore \quad R_n^\circ(\varphi_i \circ A) \leq R_n^\circ(A)$$

Q.2 Binary classification with <u>surrogate</u> loss function

$(X, P, \operatorname{sgn}(\mathcal{F}), l_{0-1})$ $\qquad Z_i = (X_i, Y_i)$ iid P

$$g_f(x) = \operatorname{sgn} f(x) = \begin{cases} +1 & \text{if } f(x) \geq 0 \\ -1 & \text{if } f(x) < 0 \end{cases}$$

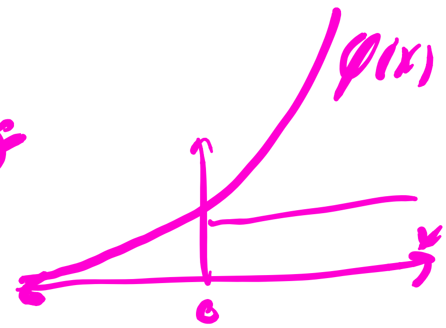
Loss using g_f on a fresh sample $(x, y) \sim_P Z$.

$$L_P(g_f) = P(g_f(x) \neq Y) = P(Y g_f(x) \leq 0)$$
$$\leq P(Y f(x) \leq 0) = E[1_{\{-Yf(x) \geq 0\}}]$$

$$\le E[\varphi(-Y f(X))]$$

where φ satisfies: $\varphi : \mathbb{R} \to \mathbb{R}$

1) φ - continuous
2) nondecreasing
3) $\varphi(x) \ge 1_{[x \ge 0]}$



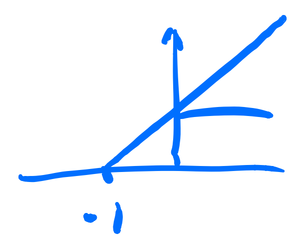
$l_\varphi(y,u) = \varphi(-yu)$

$l_\varphi(y,u)$

e^{-yu}

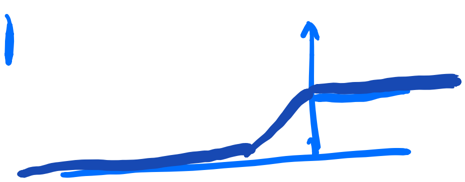
exponential $\varphi(x) = e^x$

logit $\log_2(1+e^x)$ $M_\varphi = \frac{1}{\ln 2}$

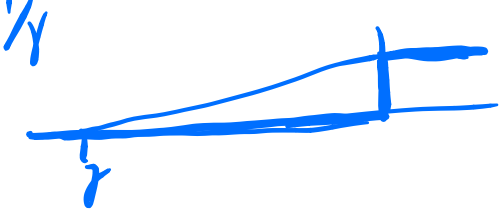


hinge $(1+x)_+$ $M_\varphi = 1$

ramp $\min[1,(1+x)_+]$ $M_\varphi = 1$



$\varphi(x/\gamma)$ $M_\varphi = 1/\gamma$



<u>Define</u> (fix $\mathbb{P}$)

$$\left\{\begin{array}{l} A_\varphi(f) = E_{\mathbb{P}}[\varphi(-Yf(X))] - \varphi(0) \quad \geq L_{\mathbb{P}}(g_f) \\ A_{\varphi,n}(f) = \frac{1}{n}\sum_{i=1}^{n}\varphi(-Y_i f(X_i)) - \varphi(0) \end{array}\right.$$

$F(x) = \varphi(x) - \varphi(0)$ — has same Lipschitz constant as φ.

<u>Lemma 8.2</u> Let φ satisfy 1–3 and suppose $\varphi(-yf(x)) \leq B$ for all x, y

and φ – Lip(M_φ)

<u>Then</u> $\Delta_n(Z^n) \left(= \sup_f |A_\varphi(f) - A_{\varphi,n}(f)|\right)$

$$\leq 4M_\varphi E R_n(\mathcal{F}(x^n)) + \frac{Bt}{\sqrt{n}}$$

with probability at least $1 - e^{-2t^2}$

$$\sqrt{\frac{2\log(1/\delta)}{n}} \qquad \delta$$