Jump back to Chapter 2

Sect. 2.5 Subgaussian r.v.

Def X is subgaussian (ν) if
$$E\left[e^{s(X-E(X))}\right] \le e^{s^2\nu^2/2} \quad \text{for all } s \in \mathbb{R}$$

$$\Rightarrow \quad P\{X - E(X) \ge \nu t\} \le e^{-t^2/2}$$

Maximal Lemma Suppose $X_1, \dots, X_n$ are $\underbrace{}_{\text{mean } 0}$ subgaussian (ν). Then

(2.31) $$\boxed{P\{\max_i X_i \ge \nu(\sqrt{2\log n} + t)\} \le e^{-t^2/2 - t\sqrt{2\log n}}}$$

Proof Union bound

$$\le \sum_i P\{X_i \ge \nu(\sqrt{2\log n} + t)\} \le n e^{-(\sqrt{2\log n} + t)^2}$$

$$P(\min_j X_i \le -\nu(\quad)) \le \cdots$$

(2.30) $$\boxed{E[\max_i X_i] \le \nu\sqrt{2\log n}}$$

Jensen's

Proof
$$e^{sE[\max_i X_i]} \le E\left[e^{s\max(X_i)}\right] = E\left[\max_i e^{sX_i}\right] \le \sum_j e^{sX_j} \le n e^{\frac{s^2\nu^2}{2}}$$

$$s = \frac{\sqrt{2\log n}}{\nu}$$

$$= \nu\sqrt{2\log n}.$$

Return to Chapter 6

(Most) abstract framework

$(\mathcal{Z}, \mathcal{P}, \mathcal{F})$

$Z^n \sim (Z_1, \cdots, Z_n)$ iid P

Find f to minimize $P(f) = E[f(Z)]$

ERM $\hat{f} = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \, P_n(f)$

$P_n(f) = \frac{1}{n} \sum_i f(Z_i)$

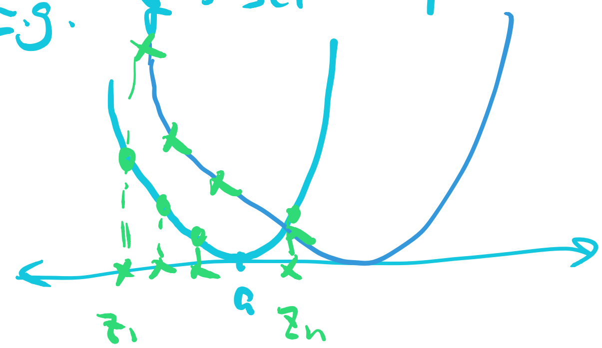
$\Delta_n(Z^n) = \sup_f |P_n(f) - P(f)|$ MMML

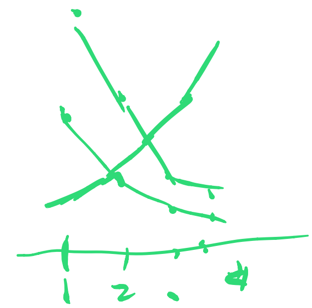
$\left(P(\hat{f}_n) \leq P^*(\mathcal{F}) + 2\Delta_n(Z^n) \right)$

Thm 6.1 $E\Delta_n(Z^n) \leq 2 E R_n(\mathcal{F}(Z^n))$

$\mathcal{F}(Z^n) = \{ (f(Z_1), \cdots, f(Z_n)) : f \in \mathcal{F} \} \subset \mathbb{R}^n$

Eg. $\mathcal{F}$ = set of paraboles $(x-a)^2$ $a \in \mathbb{R}$





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

$\mathcal{A} \in \mathbb{R}^n$

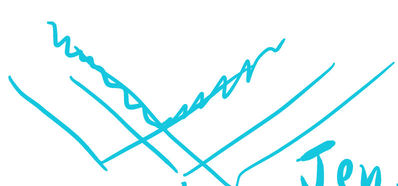
VC dimension

Thm 6.1 $\boxed{E\,\Delta_n(Z^n) \le 2\,E\,R_n(\mathcal{F}(Z^n))}$

Proof Fix $P \in \mathcal{P}$

$$E[\Delta_n(Z^n)] = E\left[\sup_f \left|P_n(f) - P(f)\right|\right]$$

convex in this argument



so by Jensen's inequality

identical, indep. copy of $P_n(f)$

$$E[\Delta_n(Z^n)] \le E\left[\sup_f \left|P_n(f) - \tilde{P}_n(f)\right|\right]$$

$\frac{1}{n}\sum_{i=1}^{n} f(Z_i)$ $\quad$ $\frac{1}{n}\sum_{i=1}^{n} f(\tilde{Z}_i)$

$$= E\left[\sup_f \frac{1}{n}\left|\sum_{i=1}^{n} \underbrace{\left(f(Z_i) - f(\bar{Z}_i)\right)}_{\text{distn is symmetric}} \varepsilon_i\right|\right]$$

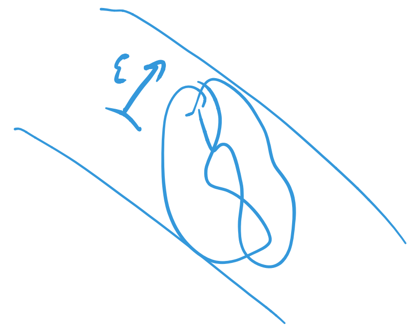
$\varepsilon_i = \pm 1$

$$\leq E\left[\sup_f \frac{1}{n}\left|\sum_{i=1}^{n} f(Z_i)\varepsilon_i\right|\right] + E\left[\sup_f \frac{1}{n}\left|\sum_{i=1}^{n} f(\bar{Z}_i)\varepsilon_i\right|\right]$$

$$= 2 E R_n(\mathcal{F}(Z^n))$$

Properties of $R_n(A)$, $R_n^0(A)$

$$R_n^0(A) \le R_n(A) = R_n^0(A \cup -A)$$



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

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

$$R_n^0(A + c) = R_n^0(A)$$

$$R_n(A \cup B) \le R_n(A) \cup R_n(B)$$

$$= \frac{1}{n} E_\varepsilon \sup_{a \in A \cup B} \left| \sum_{i=1}^n a_i \varepsilon_i \right|$$

$$\le \frac{1}{n} E_\varepsilon \left\{ \sup_{a \in A} \left| \sum_{i=1}^n a_i \varepsilon_i \right| + \sup_{a \in B} \left| \sum_{i=1}^n a_i \varepsilon_i \right| \right\}$$

$$R_n(A+B) = R_n(\{a+b : a \in A, b \in B\}) \le R_n(A) + R_n(B)$$

$$= \frac{1}{n} E_\varepsilon \sup_{a \in A, b \in B} \left| \sum_{i=1}^n (a_i + b_i) \varepsilon_i \right|$$

$$\le \frac{1}{n} E_\varepsilon \sup_{a} \left| \sum_{i=1}^n a_i \varepsilon_i \right| + \frac{1}{n} E_\varepsilon \sup_{b} \left| \sum a_i \varepsilon_i \right|$$

$$R_n(cA) = c R_n(A)$$

$\mathrm{conv}(A)$ = convex hull of A = smallest convex set containing A.

$$R_n(\mathrm{conv}(A)) = R_n(A)$$

$$\mathrm{absconv}(A) = \left\{ \sum_{m=1}^{N} c_m a^{(m)}, \; N \geq 1, \; |c_1| + \cdots + |c_N| = 1, \; a^{(1)}, \ldots, a^{(N)} \in A \right\}$$
$$= \mathrm{conv}(A \cup -A)$$

$$R_n(\mathrm{absconv}(A)) = R_n(A)$$
$$\| \qquad\qquad \|$$
$$R_n(\mathrm{conv}(A \cup A)) = R_n(A \cup A)$$

Finite Class Lemma (for Rademacher averages)

If $A = \{a^{(1)}, \ldots, a^{(N)}\} \in \mathbb{R}^n$ with

$\|a^{(j)}\|_2 \leq L$ for $1 \leq j \leq N$ then

$$R_n(A) \leq \frac{2L\sqrt{\log N}}{n}$$

Proof (Use maximal inequality for subgaussian + Hoeffding)

$$R_n(A) = E_\varepsilon \max_{1 \leq k \leq N} \Big| \underbrace{\frac{1}{n} \sum \varepsilon_i a_i^{(k)}}_{Y_k} \Big| \qquad \langle \varepsilon, a^{(k)} \rangle$$

$$= E_\varepsilon \left[\max\{ Y_1, -Y_1, Y_2, -Y_2, \ldots, Y_N, -Y_N \} \right]$$

Y_k is ν-subgaussian for $\nu^2 = \frac{1}{n^2} \sum_{i=1}^{n} (a_i^{(k)})^2 \leq \frac{L^2}{n^2}$

or $\nu = \frac{L}{n}$

$= E \{ \text{max of } 2N \text{ rv's sub. gaussian } (\frac{L}{n}) \}$

$\leq \frac{L}{n} \sqrt{2 \log 2N} \leq \frac{2L \sqrt{\log N}}{n}$ $N \geq 2$