Homework 2 Solution.

Problem 1.

First state and prove the Massart's finite Lemma:

Let $A \subseteq \mathbb{R}^n$ be a finite set with $r = \max_{a \in A} \|a\|_2$ and ϵ_i denotes the iid. Rademacher r.v., then

$$E_\epsilon \left[\max_{a \in A} \sum_{i=1}^n \epsilon_i a_i \right] \le r \sqrt{2 \log |A|}$$

where $|A|$ denotes the cardinality of set A.

Pf: for any $t > 0$,

$$\exp\left(t\, E_\epsilon \left[\max_{a \in A} \sum_{i=1}^n \epsilon_i a_i \right] \right) \overset{\text{Jensen's ineq}}{\le} E_\epsilon \left[\exp\left(t \max_{a \in A} \sum_{i=1}^n \epsilon_i a_i \right) \right]$$

exp. is monotonic

$$= E_\epsilon \left[\max_{a \in A} \exp\left(t \sum_{i=1}^n \epsilon_i a_i \right) \right]$$

ϵ_i indep.

$$\le \sum_{a \in A} E_\epsilon \left[\prod_{i=1}^n \exp(t \epsilon_i a_i) \right] = \sum_{a \in A} \prod_{i=1}^n E_{\epsilon_i} \left(\exp(t \epsilon_i a_i) \right).$$

$$\cancel{\le \sum_{a \in A} \prod_{i=1}^n \exp\left(\frac{t^2}{2}\right.} \quad = \sum_{a \in A} \prod_{i=1}^n \frac{e^{t a_i} + e^{-t a_i}}{2}.$$

by $\frac{e^x + e^{-x}}{2} \le e^{x^2/2}$,

$$\le \sum_{a \in A} \prod_{i=1}^n e^{t^2 a_i^2/2} = \sum_{a \in A} e^{t^2 \|a\|_2^2/2}.$$

$$\le \sum_{a \in A} e^{t^2 r^2/2} = |A| e^{t^2 r^2/2}.$$

Now taking the logarithm on both sides and dividing by t,

$$E\left[\max_{a \in A} \sum_{i=1}^{n} \epsilon_i a_i\right] \leq \frac{\log |A|}{t} + \frac{t r^2}{2}.$$

Take $t = \sqrt{2\log|A|/r^2}$, we obtain

$$E\left[\max_{a \in A} \sum_{i=1}^{n} \epsilon_i a_i\right] \leq r\sqrt{2\log|A|}.$$

Now consider

$$R(\mathcal{F}(X_1^n)/n) = E_\epsilon\left[\sup_{f \in \mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(x_i)\right|\right]$$

$$= E_\epsilon\left[\sup_{f \in \mathcal{F}} \left(\max\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(x_i), \frac{1}{n}\sum_{i=1}^{n}\epsilon_i(-f(x_i))\right)\right)\right]$$

$$= E_\epsilon \sup_{f \in \mathcal{F} \cup \mathcal{F}^-} \left(\frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(x_i)\right)$$

define $\mathcal{F}^- = \{-f : \forall f \in \mathcal{F}\}$. and it's easy to see $|\mathcal{F}| = |\mathcal{F}^-|$

and $\sup_{f \in \mathcal{F}} \sqrt{\frac{\sum_{i=1}^{n} f^2(x_i)}{n}} = \sup_{f \in \mathcal{F}^-} \sqrt{\frac{\sum_{i=1}^{n} f^2(x_i)}{n}}$

Then apply Massart's lemma,

$$= \sup_{f \in \mathcal{F} \cup \mathcal{F}} \sqrt{\frac{\sum_{i=1}^{n} f^2(x_i)}{n}}$$

$$=: D(X_1^n)$$

$$R(\mathcal{F}(X_1^n)/n) \leq D(X_1^n) \cdot \sqrt{\frac{2\log \Pi_{\mathcal{F} \cup \mathcal{F}^-}}{n}}$$

$$= D(X_1^n)\sqrt{\frac{2\log(2\Pi_{\mathcal{F}})}{n}}.$$

□

Problem 2.

By def. $\mathcal{G}(\phi(T)) = E[\sup_{\theta\in T} \sum g_i \phi(\theta_i)]$

$$\mathcal{G}(T) = E[\sup_{\theta\in T} \sum g_i\theta_i]$$

Applying Sudakov - Fernique thm,

let $X_\theta = \langle g, \phi(\theta)\rangle$ and $Y_\theta = \langle g, \theta\rangle$.

Now to show for ~~and~~ any $\theta, \theta' \in T$,

$$E(X_\theta - X_{\theta'})^2 \leq E(Y_\theta - Y_{\theta'})^2$$

pf : $E(X_\theta - X_{\theta'})^2 = E(\langle g, \phi(\theta)-\phi(\theta')\rangle)^2$

$$= E\left(\sum_{i=1} g_i (\phi(\theta_i)-\phi(\theta_i'))\right)^2$$

by g_i indep. $= \sum_{i=1} E(g_i)^2 \cdot (\phi(\theta_i) - \phi(\theta_i'))^2$

since ϕ is 1-Lipschitz fn.

$$\leq \sum E(g_i)^2 \cdot (\theta_i - \theta_i')^2 = E(Y_\theta - Y_{\theta'})^2$$

~~Thus~~ And easy to see that $EX_\theta = EY_\theta = 0$

Then $E[\sup_{\theta\in T} X_\theta] \leq E[\sup_{\theta\in T} Y_\theta]$

$$\Rightarrow \mathcal{G}(\phi(T)) \leq \mathcal{G}(T) \qquad \square$$

Problem 3.

① First to show $R(T) \le \sqrt{\frac{\pi}{2}}\, \mathcal{G}(T)$.

Note that $g \sim N(0,1)$, then $E|g| = \sqrt{\frac{2}{\pi}}$

$$R(T) = E_\epsilon\left[\sup_{\theta\in T} \sum_{i=1}^d \theta_i \epsilon_i\right] = \sqrt{\frac{\pi}{2}}\, E_\epsilon\left[\sup_{\theta\in T} \sum_{i=1}^d \theta_i \epsilon_i E|g_i|\right]$$

$$= \sqrt{\frac{\pi}{2}}\, E_\epsilon\left[\sup_{\theta\in T} E_g\left(\sum_{i=1}^d \epsilon_i |g_i| \cdot \theta_i\right)\right].$$

$$\le \sqrt{\frac{\pi}{2}}\, E_\epsilon E_g\left[\sup_{\theta\in T} \sum_{i=1}^d \epsilon_i |g_i| \cdot \theta_i\right] \quad \text{by Jensen's ineq.}$$

Note that $\epsilon_i |g_i| = \epsilon_i \cdot \mathrm{Sign}(g_i) \cdot g_i \overset{d}{=} g_i$ (where $\epsilon_i \cdot \mathrm{Sign}(g_i) \overset{d}{=} \epsilon_i$)

$$\le \sqrt{\frac{\pi}{2}}\, E_g\left[\sup_{\theta\in T} \sum_{i=1}^d g_i \theta_i\right] = \sqrt{\frac{\pi}{2}}\, \mathcal{G}(T).$$

② now show $\sqrt{\frac{\pi}{2}}\, \mathcal{G}(T) \le c\sqrt{\log d}\, R(T)$ for $c>0$ universal const.

$(\Leftrightarrow)$ $\mathcal{G}(T) \le c'\sqrt{\log d} \cdot R(T)$ for some $c'>0$.

$$E_g \sup_{\theta\in T}\left[\sum_{i=1}^d g_i\theta_i\right] = E_g \sup_{\theta\in T}\left[\sum_{i=1}^d \epsilon_i\theta_i |g_i|\right]$$

$$\le E_g\left(\max_i |g_i|\right) \cdot E_\epsilon\left[\sup_{\theta\in T} \sum_{i=1}^d \epsilon_i\theta_i\right]$$

$$\le \sqrt{2\log d} \cdot R(T) = c'\sqrt{\log d}\, R(T). \quad \square$$

Problem 4.

(a) Consider the set $S^d(s) = \{ \theta \in \mathbb{R}^d : \|\theta\|_0 \leq s \ \|\theta\|_2 \leq 1 \}$.

It's easy to see that

$$\sup_{\|\theta\|_2 \leq 1} \langle w, \theta \rangle = \|w\|_2 \quad \text{for } w \sim N(0,1).$$

$$\sup_{\substack{\|\theta\|_0 \leq s, \\ \|\theta\|_2 \leq 1}} \langle w, \theta \rangle = \max_{|S|=s} \|w_S\|_2.$$

$w_S \in \mathbb{R}^{|S|}$ is a sub-vector of $(w_1, \cdots, w_d)$.

Now to show that $\|\cdot\|_2$ is 1-Lipschitz

for any w, w',

$$\left(\|w\|_2 - \|w'\|_2 \right)^2 = \sum w_i^2 + \sum w_i'^2 - 2\sqrt{\sum w_i^2}\sqrt{\sum w_i'^2}$$

by Cauchy-Schwartz ‡.

$$\leq \sum w_i^2 + \sum w_i'^2 - 2 \sum w_i w_i' = \|w - w'\|^2.$$

By applying the concentration ineq for Lipschitz function.

$$P\left(\|w_S\|_2 - E\|w_S\|_2 \geq t \right)$$

$$= P\left(\|w_S\|_2 \geq \sqrt{s} + t \right) \leq e^{-\frac{t^2}{2}} \Rightarrow \|w_S\|_2 \text{ is } 1\text{-subgaussian r.v.}$$

Apply the union bound.

$$P\Big[\max_{|S|=s} \|W_S\|_2 \geq \sqrt{s}+t\Big] \leq \binom{d}{s} e^{-\frac{t^2}{2}}$$

Now consider $\max_{|S|=s} \|W_S\|_2$, similar calculation as problem 1.

$$\exp\Big(t \cdot E\big(\max_{|S|=s} \|W_S\|_2\big)\Big) \overset{\text{Jensen's} \neq}{\leq} E \exp\Big(t \cdot \max_{|S|=s} \|W_S\|_2\Big)$$

$$\leq E \max_{|S|=s} \exp\big(t \cdot \|W_S\|_2\big)$$

$$\leq \sum_{|S|=s} E \exp\big(t\, \|W_S\|_2\big) \leq \binom{d}{s} \cdot e^{\frac{t^2}{2}}$$

since $\|W_S\|_2$ is 1-sub gaussian. r.v.

Now take the logarithm and divide by t on both sides.

$$E\big(\max_{|S|=s} \|W_S\|_2\big) \leq \frac{\log\binom{d}{s}}{t} + \frac{t}{2}$$

Take $t = \sqrt{2\log\binom{d}{s}}$, and note that $\binom{d}{s} = \frac{d!}{s!(d-s)!} \leq \frac{d^s}{s!} \leq \frac{(de)^s}{s^s}$.

$s! \geq s^s e^{-s}$

$$E\big(\max_{|S|=s} \|W_S\|_2\big) \leq c\sqrt{s \cdot \log\left(\frac{ed}{s}\right)}$$ $\square$

Problem 4

① upper bound of $\mathcal{G}(B_q^d(1))$

(b) $$\mathcal{G}(B_q^d(1)) = E\left(\sup_{\|\theta\|_q \le 1} \langle g, \theta \rangle\right)$$

$$= E\left(\sup_{\|\theta\|_q \le 1} \sum_{i=1}^{d} g_i \theta_i\right) = E\left(\|g\|_*\right)$$

define $\|\cdot\|_*$ is the dual norm,

By the fact that :

$$\sup\left\{\sum_{i=1}^{d} g_i \theta_i : \theta \in \mathbb{R}^d, \|\theta\|_q \le 1\right\} = \|z\|_p$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

$$\Rightarrow \quad E\left(\|g\|_*\right) = E\left(\left(\sum_{i=1}^{d} |g_i|^p\right)^{\frac{1}{p}}\right)$$

$$\le \left(E \sum_{i=1}^{d} |g_i|^p\right)^{\frac{1}{p}} \quad \text{by Jensen's } \neq$$

$$= \left(d \cdot c_0\right)^{\frac{1}{p}} \quad \text{define } c_0 = E\left(|g_i|^p\right)$$

$$\Rightarrow \mathcal{G}(B_q^d(1)) \le d^{\frac{1}{p}} c_0^{\frac{1}{p}} = d^{1-\frac{1}{q}} \cdot c_q .$$

②. lower bound of $\mathcal{G}(B_q^d(1))$.

By Hölder's $\neq$, for $\frac{1}{p}+\frac{1}{q}=1$

$$\sum_{i=1}^{d}(|x_i|\cdot 1) \le \left(\sum_{i=1}^{d}|x_i|^p\right)^{\frac{1}{p}} \cdot \underbrace{\left(\sum_{i=1}^{d} 1^q\right)^{\frac{1}{q}}}_{d^{\frac{1}{q}}.}$$

$$\Rightarrow \quad E\left[\sum_{i=1}^{d}|g_i|\right] \le E\left[\left(\sum_{i=1}^{d}|g_i|^p\right)^{\frac{1}{p}}\right] \cdot d^{\frac{1}{q}}$$

$$= \mathcal{G}\left(B_q^d(1)\right) \cdot d^{\frac{1}{q}}$$

Since.

$$E|g_i| = \sqrt{\tfrac{2}{\pi}}$$

$$\Rightarrow \quad d\cdot\sqrt{\tfrac{2}{\pi}} \le \mathcal{G}\left(B_q^d(1)\right) d^{\frac{1}{q}}.$$

$$\Rightarrow \quad d^{1-\frac{1}{q}}\sqrt{\tfrac{2}{\pi}} \le \mathcal{G}\left(B_q^d(1)\right)$$

$\square$

Problem 5,

First show for finite case. Suppose $|T| = n$.

Then for $\{X_1, \cdots, X_n\}$. we can write

$$X_i = a_{i1} g_1 + a_{i2} g_2 + \cdots + a_{in} g_n \quad \text{for } i = 1, \cdots, n.$$

and $g_i \overset{iid}{\sim} N(0,1)$

Define the matrix $A = (a_{ij})_{n \times n}$. and $g = (g_1, \cdots, g_n)$

Then $Ag \overset{d}{=} (X_1, \cdots, X_n)$.

Consider function $F(x) = \max \{ (Ax)_i : i = 1, \cdots, n \}$

It is easy to see that $\forall x, x'$

$$\begin{aligned} |F(x) - F(x')| &= |\max (Ax)_i - \max (Ax')_i| \\ &\le \max |A(x-x')_i| \le \max_i \sqrt{\sum_j a_{ij}^2} \, \|x - x'\|_2. \\ &= \max_i \sqrt{EX_i^2} \, \|x - x'\|_2 \end{aligned}$$

Then apply the Concentration inequalities for Lipschitz function.

$$P\left(|\max_i X_i - E \max_i X_i| > t \right) \le 2 \exp\left(\frac{-t^2}{\max_i E(X_i^2)} \right).$$

Now to extend to the infinit case.

for any fixed n, define iid random copy Y_i of X_i, $i=1,\cdots,n$.

Then consider large ~~fixe~~ enough fixed n, for $m \leq n$,

$$\left| \max_{1\le i\le n} X_i - E \max_{1\le i \le n} X_i \right|$$

$$= \left| \max_{1\le i\le n} X_i - E\max_{1\le j\le n} Y_j + E\sup_{\theta\in T} X_\theta - E \sup_{\theta\in T} X_\theta \right|$$

$$\le \left| \max_{1\le i\le n} X_i - E\sup_{\theta\in T} X_\theta \right| + \left| E\sup_{\theta\in T} X_\theta - E\max_{1\le j\le n} Y_j \right|$$

$$\le \left| \max_{1\le i\le n} X_i - E\sup_{\theta\in T} X_\theta \right| + \underbrace{\left| E\sup_{\theta\in T} X_\theta - E\max_{1\le j\le m} Y_j \right|}_{=\epsilon_m}.$$

~~for fixed m,~~ choose m large enough s.t $t-\epsilon_m > 0$.

Now define the seq of sets $A_{nm} = \left\{ \left|\max_{i\le n} X_i - E\sup_{\theta\in T} X_\theta \right| > t-\epsilon_m \right\}$.

for any fixed m, $A_{n,m} \subset A_{n+1,m}$, and since T is a countible set

and $\bigcup_{n=1}^{\infty} A_{n,m} = \lim_{n\to\infty} A_{n,m} = \left\{ \left| \sup_{\theta\in T} X_\theta - E\sup_{\theta\in T} X_\theta \right| > t-\epsilon_m \right\}$.

Then we have $p(\lim_{n\to\infty} A_{n,m}) = \lim_{n\to\infty} p(A_{n,m})$.

and $A_{\infty,m} \supset A_{\infty,m+1}$ since $\epsilon_m \downarrow 0$ as $m\to\infty$. similarly,

$$p(\lim_{n,m\to\infty} A_{n,m}) = \lim_{n,m\to\infty} p(A_{n,m}) \le \lim_{n,m\to\infty} e^{-\frac{(t-\epsilon_m)^2}{2\max_{i\le n} EX_i^2}} = e^{-\frac{t^2}{2(\sup_\theta EX_\theta^2)}}$$

$\square$