Homework 1

1. Let H be a Hilbert space, and G be a bounded set in H, i.e. there exists $b > 0$ such that $\|g\| \le b$ holds for any $g \in G$. Let $f \in \overline{Conv(G)}$. Prove that for any integer $m > 0$, there exist $g_1, g_2, ..., g_m \in G$, $\gamma_1, \gamma_2, ..., \gamma_m \ge 0$ that satisfies $\sum_{i=1}^m \gamma_i = 1$, and

$$\|f - \sum_{i=1}^m \gamma_i g_i\| \le \frac{b}{\sqrt{m}}.$$

2. **Rademacher complexity**

Recall that for a function class $\mathcal{F}$ and data $\{x_i\}_{i=1}^n$, we use $\mathrm{Rad}_n(\mathcal{F})$ to denote the empirical Rademacher complexity of $\mathcal{F}$ at $\{x_i\}_{i=1}^n$:

$$\mathrm{Rad}_n(\mathcal{F}) = \frac{1}{n}\mathbb{E}_\sigma \sup_{f\in\mathcal{F}} \sum_{i=1}^n \sigma_i f(x_i).$$

a. Let $\{x_i\}_{i=1}^n$ be n points in $\mathbb{R}^d$ that satisfies $\|x_i\|_\infty \le 1$ for any $i = 1, 2, ..., n$. Show that

$$\frac{1}{n}\mathbb{E}_\sigma\left[\|\sum_{i=1}^n \sigma_i x_i\|_\infty\right] \le \sqrt{\frac{2\log d}{n}}.$$

(Hint: use the Massart lemma.)

b. Let $\mathcal{F}$ be the class of linear predictors with the L_1-norm of the weights bounded by W, i.e. $\mathcal{F} = \{w^T x : \|w\|_1 \le W\}$. Show that

$$\mathrm{Rad}_n(\mathcal{F}) \le \max_{1\le i\le n} \|x_i\|_\infty W\sqrt{\frac{2\log d}{n}}.$$

c. Assume $\mathcal{D}$ is a distribution on $\mathbb{R}^d$ and all x sampled from $\mathcal{D}$ satisfies $\|x\|_\infty \le 1$. Let $S = \{x_1, x_2, ..., x_n\}$ be n points i.i.d. sampled from $\mathcal{D}$. For any $f \in \mathcal{F}$, write down the upper bound of the generalization gap $\mathbb{E}_\mathcal{D}[f(x)] - \hat{\mathbb{E}}_S[f(x)]$ using the bound derived in (b).

d. Show that, there exists a constant c, such that for any $\delta \in (0, 1)$, with probability no less than $1 - \delta$ over the choice of S, we have

$$\mathbb{E}_\mathcal{D}[f(x)] \le \hat{\mathbb{E}}_S[f(x)] + (\|w\|_1 + 1)\sqrt{\frac{2\log d}{n}} + 3\sqrt{\frac{2\log(c(\|w\|_1+1)^2)/\delta}{n}},$$

for any linear predictor $f(x) = w^T x$ (with no constraint on $\|w\|_1$).

3. **Covering and Packing number**

a. Let B be the unit ball in $\mathbb{R}^d$: $B = \{x \in \mathbb{R}^d : \|x\|_2 \le 1\}$. Show that

$$N(\alpha, B, l_2) \ge \left(\frac{1}{\alpha}\right)^d.$$

b. For a metric space (S, d) and a set $T \subset S$, we say T' is an α-packing of T if $T' \subset T$ and for any $x_1, x_2 \in T'$, there is $d(x_1, x_2) > \alpha$. Let $M(\alpha, T, d)$ be the packing number of T, which is defined as

$$M(\alpha, T, d) = \max\{|T'| : T'\ is\ an\ \alpha - packing\ of\ T\}.$$

Prove that

$$M(2\alpha, B, l_2) \leq N(\alpha, B, l_2) \leq M(\alpha, B, l_2).$$

c. Show that

$$N(\alpha, B, l_2) \leq \left(\frac{2}{\alpha} + 1\right)^d.$$