Theory of Machine Learning: Homework 3

1. We want to show that $A \subseteq B$ does not imply $N(A, \epsilon) \le N(B, \epsilon)$.

   Proof. We can provide a counterexample. Consider $A \subset \mathbb{R}$, $\epsilon > 0$. Let $A = \{0, \frac{3}{2}\epsilon\}$. The smallest ϵ-net that will cover A is itself, i.e. we need a ball of radius ϵ at both points in A to cover it, since they are farther than ϵ away from one another and cannot both be covered by a single ball. So $N(A, \epsilon) = 2$.

   Create $B = A \cup \{\frac{3}{4}\epsilon\} = \{0, \frac{3}{4}\epsilon, \frac{3}{2}\epsilon\}$. Note that $|\frac{3}{4}\epsilon - 0| = \frac{3}{4}\epsilon$ and $|\frac{3}{4}\epsilon - \frac{3}{2}\epsilon| = \frac{3}{4}\epsilon$, so a single ball of radius ϵ placed at the new point, $\frac{3}{4}\epsilon$, will cover all of B. So smallest ϵ-net over B is $\{\frac{3}{4}\epsilon\} \implies N(B, \epsilon) = 1$.

   So it is simultaneously true that $A \subseteq B$ and $N(A, \epsilon) > N(B, \epsilon)$. QED

2. *N/A*

3. (a) We want to prove for any S and any $h, h' \in \mathcal{H}$ that $|E_S(h) - E_S(h')| \le 4M \|h - h'\|_\infty$.

   Proof.

$$
\begin{aligned}
|E_S(h) - E_S(h')| &= \left|[L_D(h) - \hat{L_S}(h)] - [L_D(h') - \hat{L_S}(h')]\right| \\
&= \left|[L_D(h) - L_D(h')] + [\hat{L_S}(h') - \hat{L_S}(h)]\right| \\
&\le |[L_D(h) - L_D(h')]| + \left|[\hat{L_S}(h') - \hat{L_S}(h)]\right| \quad \text{(triangle inequality)} \\
&= \left|\mathbb{E}_{z\sim D}\left[(h(x) - y)^2 - (h'(x) - y)^2\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}\left[(h'(x_i) - y_i)^2 - (h(x_i) - y_i)^2\right]\right| \\
&= \left|\mathbb{E}_{z\sim D}\left[h(x)^2 - 2h(x)y + y^2 - (h'(x)^2 - 2h'(x)y + y^2)\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \\
&= \left|\mathbb{E}_{z\sim D}\left[(h(x)^2 - h'(x)^2) - 2y(h'(x) - h(x))\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \\
&= \left|\mathbb{E}_{z\sim D}\left[(h(x) + h'(x))(h(x) - h'(x)) - 2y(h'(x) - h(x))\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \\
&= \left|\mathbb{E}_{z\sim D}\left[(h(x) - h'(x))(h(x) + h'(x) - 2y)\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \\
&\le \left|\mathbb{E}_{z\sim D}\left[\max_{x\in\mathcal{X}}[h(x) - h'(x)] \cdot [(h(x) - y) + (h'(x) - y)]\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \quad \text{(monotonicity)} \\
&\le \left|\mathbb{E}_{z\sim D}\left[\|h - h'\|_\infty \cdot (M + M)\right]\right| + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \quad \text{(monotonicity)} \\
&= 2M\|h - h'\|_\infty + \left|\frac{1}{m}\sum_{i=1}^{m}[\cdots]\right| \\
&\le 2M\|h - h'\|_\infty + \left|\frac{1}{m}\sum_{i=1}^{m}2M\|h - h'\|_\infty\right|, \quad \text{by the same process as above} \\
&= 4M\|h - h'\|_\infty \\
\implies |E_S(h) - E_S(h')| &\le 4M\|h - h'\|_\infty
\end{aligned}
$$

   QED

   (b) Assume $\mathcal{H} \subset \bigcup_{j=1}^{k} \mathcal{B}_j$. We want to show that

$$
\Pr_{S\sim D^m}\left[\sup_{h\in\mathcal{H}}|E_S(h)| \ge \epsilon\right] \le \sum_{j=1}^{k}\Pr_{S\sim D^m}\left[\sup_{h\in\mathcal{B}_j}|E_S(h)| \ge \epsilon\right]
$$

Proof. Define $E = \{S : \sup_{h\in\mathcal{H}} |E_S(h)| \geq \epsilon\}$. Define $E_j = \{S : \sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\}$. Take any $S' \in E$. A sequence h_n must exist s.t. $\lim_{n\to\infty} |E_{S'}(h_n)| \to \sup_{h\in\mathcal{H}} |E_{S'}(h_n)|$. There are k sets covering $\mathcal{H}$. So, for at least one index j', there should be a subsequence $h_{n_l} \in \mathcal{B}_{j'}$, i.e. there are infinitely many points occuring in one set $\mathcal{B}_{j'}$. So,

$$\sup_{h\in\mathcal{B}_{j'}} |E_{S'}(h)| \geq \lim_{l\to\infty} |E_{S'}(h_{n_l})| = \sup_{h\in\mathcal{H}} |E_{S'}(h)| \geq \epsilon$$

That is, for all $S \in E$, for some $j \in [k]$, $S \in E_j$. So, $E \subseteq \bigcup_{j=1}^k E_j$. So,

$$D^m\left(\{S : \sup_{h\in\mathcal{H}} |E_S(h)| \geq \epsilon\}\right) \leq D^m\left(\bigcup_{j=1}^{k}\{S : \sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\}\right) \text{ (monotonicity of } D^m)$$

$$\implies D^m\left(\{S : \sup_{h\in\mathcal{H}} |E_S(h)| \geq \epsilon\}\right) \leq \sum_{j=1}^{k} D^m\left(\{S : \sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\}\right) \text{ (union bound)}$$

QED

(c) Let $\mathcal{B}_j$ be the intersection of $\mathcal{H}$ with a ball of radius $\epsilon/(8M)$ w.r.t. the L^∞ norm, centered at h_j. We want to show that

$$\Pr\left[\sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\right] \leq \Pr\left[|E_S(h_j)| \geq \epsilon/2\right]$$

Proof. Define $E = \{S : \sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\}$. Define $E' = \{S : |E_S(h_j)| \geq \epsilon/2\}$. We know that $\forall h \in \mathcal{B}_j$, $\|h - h_j\|_\infty \leq \epsilon/(8M)$. Then

$$\begin{aligned}|E_S(h) - E_S(h_j)| &\leq 4M\,\|h - h_j\|_\infty \text{ (by part (a))}\\ &\leq 4M\left(\frac{\epsilon}{8M}\right)\\ &= \epsilon/2\end{aligned}$$

Since $||E_S(h)| - |E_S(h_j)|| \leq |E_S(h) - E_S(h_j)| \leq \epsilon/2$, we know that

$$\begin{aligned}|E_S(h)| - |E_S(h_j)| &\leq \epsilon/2\\ \implies |E_S(h)| &\leq |E_S(h_j)| + \frac{\epsilon}{2}\end{aligned}$$

Now, pick any $S' \in E$. It is true that

$$\begin{aligned}\epsilon \leq |E_{S'}(h)| &\leq |E_{S'}(h_j)| + \frac{\epsilon}{2}\\ \implies \epsilon &\leq |E_{S'}(h_j)| + \frac{\epsilon}{2}\\ \implies \epsilon/2 &\leq |E_{S'}(h_j)|\end{aligned}$$

So, $E \subseteq E'$. Then, the monotonicity of D^m implies

$$D^m\left(\{S : \sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\}\right) \leq D^m\left(\{S : |E_S(h_j)| \geq \epsilon/2\}\right)$$

QED

(d) Let $k = N(\mathcal{H}, \|\cdot\|_\infty, \epsilon/(8M))$ be the covering number of $\mathcal{H}$ by an ϵ-net $\{h_1, \ldots, h_k\} \subset \mathcal{H}$. We will show a generalization bound using this net.

Proof. For every $i \in [m]$, define the r.v. $X_i = \frac{1}{m}(h_j(x_i) - y_i)^2$. Define $A_m = \sum_{i=1}^m X_i$.

Note that $|h_j(x) - y| \leq M \implies X_i \in \left[0, \frac{M^2}{m}\right]$. Then, by Hoeffding's inequality,

$$\begin{aligned}\Pr[A_m - \mathbb{E}[A_m] \geq \epsilon/2] &\leq \exp\left(\frac{-2(\epsilon/2)^2}{\sum_{i=1}^m (M^2/m - 0)^2}\right)\\ &= \exp\left(\frac{-2\epsilon^2}{4\sum_{i=1}^m M^4/m^2}\right)\end{aligned}$$

$$= \exp\left(\frac{-\epsilon^2}{2M^4/m}\right)$$
$$= \exp\left(\frac{-m\epsilon^2}{2M^4}\right)$$

It is therefore also true that

$$\Pr[A_m - \mathbb{E}[A_m] \leq -\epsilon/2] \leq \exp\left(\frac{-m\epsilon^2}{2M^4}\right)$$

Both events are disjoint, so we can take the union and sum the probabilities:

$$\Pr[|A_m - \mathbb{E}[A_m]| \geq \epsilon/2] \leq 2\exp\left(\frac{-m\epsilon^2}{2M^4}\right)$$

Note that $A_m = \hat{L_S}(h_j)$ and $\mathbb{E}[A_m] = L_D(h_j)$. So,

$$|A_m - \mathbb{E}[A_m]| = |\mathbb{E}[A_m] - A_m| = |L_D(h_j) - \hat{L_S}(h_j)| = |E_S(h_j)|$$

Then, the prior inequality becomes:

$$\Pr[|E_S(h_j)| \geq \epsilon/2] \leq 2\exp\left(\frac{-m\epsilon^2}{2M^4}\right)$$

We know from (b) and (c) that

$$\Pr\left[\sup_{h\in\mathcal{H}} |E_S(h)| \geq \epsilon\right] \leq \sum_{j=1}^{k} \Pr\left[\sup_{h\in\mathcal{B}_j} |E_S(h)| \geq \epsilon\right] \leq \sum_{j=1}^{k} \Pr[|E_S(h_j)| \geq -\epsilon/2]$$

So, applying the result of Hoeffding's inequality, we get

$$\begin{aligned}
\Pr\left[\sup_{h\in\mathcal{H}} |E_S(h)| \geq \epsilon\right] &\leq \sum_{j=1}^{k} \Pr[|E_S(h_j)| \geq -\epsilon/2] \\
&\leq \sum_{j=1}^{k} 2\exp\left(\frac{-m\epsilon^2}{2M^4}\right) \\
&= k \cdot 2\exp\left(\frac{-m\epsilon^2}{2M^4}\right) \\
\implies \Pr\left[\sup_{h\in\mathcal{H}} |E_S(h)| \geq \epsilon\right] &\leq N(\mathcal{H}, \|\cdot\|_\infty, \epsilon/(8M)) \cdot 2\exp\left(\frac{-m\epsilon^2}{2M^4}\right)
\end{aligned}$$

QED