Theory of Machine Learning: Homework 3

1. Let $\mathcal{H} = \{\mathbf{x} \to \langle \mathbf{w}, \mathbf{x} \rangle$ s.t. $\|w\|_2 \leq 1\}$ be linear classifiers. Let $S_x = (\mathbf{x}_1, \dots, \mathbf{x}_m) \subset \mathbb{R}^n$ be a training set with Frobenius norm $\|S_x\|_F^2 = \sum_{i=1}^m \|\mathbf{x}_i\|_2^2$. Define $\mathcal{H} \circ S_x = \{(a_1, \dots, a_m)$ s.t. $h \in \mathcal{H},\ a_i = h(\mathbf{x}_i)\}$.

   (a) We want to show that $\hat{\mathcal{R}}_S(\mathcal{H} \circ S_x) \leq \frac{1}{m}\mathbb{E}_\sigma \left[\|\sum_{i=1}^m \sigma_i \mathbf{x}_i\|_2\right]$.

   Proof. We first want to double check that an inequality we use frequently is correct. Suppose

$$\forall t,\ f(t) \leq g(t) \implies \forall t,\ f(t) \leq g(t) \leq \sup_t g(t) \implies \sup_t f(t) \leq \sup_t g(t)$$

   The empirical Rademacher complexity is given by

$$\hat{\mathcal{R}}_S(F) = \mathbb{E}_\sigma \left[\sup_{f \in F} \frac{1}{m} \sum_{i=1}^m \sigma_i f(z_i)\right].$$

   With $F = \mathcal{H} \circ S_x$, we can rewrite this as:

$$\begin{aligned}
\hat{\mathcal{R}}_S(\mathcal{H} \circ S_x) &= \mathbb{E}_\sigma \left[\sup_{\|\mathbf{w}\| \leq 1} \frac{1}{m} \sum_{i=1}^m \sigma_i \langle \mathbf{w}, \mathbf{x}_i \rangle\right] \\
&= \frac{1}{m}\mathbb{E}_\sigma \left[\sup_{\|\mathbf{w}\| \leq 1} \sum_{i=1}^m \sigma_i \langle \mathbf{w}, \mathbf{x}_i \rangle\right] \\
&= \frac{1}{m}\mathbb{E}_\sigma \left[\sup_{\|\mathbf{w}\| \leq 1} \left\langle \mathbf{w}, \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\rangle\right] \quad \text{(bilinearity of the inner product)} \\
&\leq \frac{1}{m}\mathbb{E}_\sigma \left[\sup_{\|\mathbf{w}\| \leq 1} \left|\left\langle \mathbf{w}, \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\rangle\right|\right] \\
&\leq \frac{1}{m}\mathbb{E}_\sigma \left[\sup_{\|\mathbf{w}\| \leq 1} \|\mathbf{w}\|_2 \left\|\sum_{i=1}^m \sigma_i \mathbf{x}_i\right\|_2\right] \quad \text{(Holder's inequality)} \\
\implies \hat{\mathcal{R}}_S(\mathcal{H} \circ S_x) &\leq \frac{1}{m}\mathbb{E}_\sigma \left[\left\|\sum_{i=1}^m \sigma_i \mathbf{x}_i\right\|_2\right]
\end{aligned}$$

   QED

   (b) We want to show that $\hat{\mathcal{R}}_S(\mathcal{H} \circ S_x) \leq \frac{1}{m}\|S_x\|_F$.

   Proof. Let $g(t) = -\sqrt{t}$.

$$g'(t) = -\frac{1}{2}t^{-1/2} \implies g''(t) = \frac{1}{4}t^{-3/2}$$

   For $t \in (0, \infty)$ it is true that $g''(t) > 0$ (g is convex). For an r.v. θ Jensen's inequality therefore says

$$g(\mathbb{E}[\theta]) \leq \mathbb{E}[g(\theta)] \implies -\sqrt{\mathbb{E}[\theta]} \leq \mathbb{E}[-\sqrt{\theta}] \implies \mathbb{E}[\sqrt{\theta}] \leq \sqrt{\mathbb{E}[\theta]}$$

   So:

$$\begin{aligned}
\hat{\mathcal{R}}_S(\mathcal{H} \circ S_x) &\leq \frac{1}{m}\mathbb{E}_\sigma \left[\left\|\sum_{i=1}^m \sigma_i \mathbf{x}_i\right\|_2\right] \quad \text{(result of (a))} \\
&= \frac{1}{m}\mathbb{E}_\sigma \left[\sqrt{\left\|\sum_{i=1}^m \sigma_i \mathbf{x}_i\right\|_2^2}\right] \\
&\leq \frac{1}{m}\sqrt{\mathbb{E}_\sigma \left[\left\|\sum_{i=1}^m \sigma_i \mathbf{x}_i\right\|_2^2\right]} \quad \text{(Jensen's inequality)} \\
&= \frac{1}{m}\sqrt{\mathbb{E}_\sigma \left[\left\langle \sum_{i=1}^m \sigma_i \mathbf{x}_i, \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\rangle\right]}
\end{aligned}$$

$$= \frac{1}{m}\sqrt{\mathbb{E}_\sigma\left[\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_i\sigma_j\left\langle \mathbf{x}_i,\, \mathbf{x}_j\right\rangle\right]} \text{ (bilinearity of the inner product)}$$

$$= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\mathbb{E}_\sigma\left[\sigma_i\sigma_j\right]\left\langle \mathbf{x}_i,\, \mathbf{x}_j\right\rangle} \text{ (linearity of the expectation)}$$

When $i = j$, $\sigma_i\sigma_j = \sigma_i^2 \implies \mathbb{E}_\sigma\left[\sigma_i\sigma_j\right] = 1$. When $i \neq j$, $\mathbb{E}_\sigma\left[\sigma_i\sigma_j\right] = \mathbb{E}[\sigma_i]\mathbb{E}[\sigma_j] = 0$ by independence. So:

$$\begin{aligned}
\hat{\mathcal{R}}_S(\mathcal{H}\circ S_x) &\leq \frac{1}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\mathbb{E}_\sigma\left[\sigma_i\sigma_j\right]\left\langle \mathbf{x}_i,\, \mathbf{x}_j\right\rangle} \\
&= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\left\langle \mathbf{x}_i,\, \mathbf{x}_i\right\rangle} \\
&= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\|\mathbf{x}_i\|_2^2} \\
\implies \hat{\mathcal{R}}_S(\mathcal{H}\circ S_x) &\leq \frac{1}{m}\|S_x\|_F
\end{aligned}$$

QED

(c) Now suppose $D \sim \mathcal{N}(0, I_{n\times n})$. We want to bound the Rademacher complexity $\mathcal{R}_m(\mathcal{H})$.

Proof.

$$\begin{aligned}
\hat{\mathcal{R}}_S(\mathcal{H}\circ S_x) &\leq \frac{1}{m}\|S_x\|_F \\
\mathbb{E}_{S_x\sim D^m}\left[\hat{\mathcal{R}}_S(\mathcal{H}\circ S_x)\right] &\leq \mathbb{E}_{S_x\sim D^m}\left[\frac{1}{m}\|S_x\|_F\right] \text{ (monotonicity of the expectation)} \\
\text{So, } \mathcal{R}_m(\mathcal{H}) &\leq \frac{1}{m}\mathbb{E}_{S_x\sim D^m}\left[\|S_x\|_F\right] \\
&= \frac{1}{m}\mathbb{E}_{S_x\sim D^m}\left[\sqrt{\sum_{i=1}^{m}\|\mathbf{x}_i\|_2^2}\right] \\
&\leq \frac{1}{m}\sqrt{\mathbb{E}_{S_x\sim D^m}\left[\sum_{i=1}^{m}\|\mathbf{x}_i\|_2^2\right]} \text{ (Jensen's inequality)} \\
&= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\mathbb{E}_{S_x\sim D^m}\left[\|\mathbf{x}_i\|_2^2\right]} \text{ (linearity of the expectation)} \\
&= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\mathbb{E}_{S_x\sim D^m}\left[\sum_{j=1}^{n}(\mathbf{x}_i)_j^2\right]} \\
&= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{E}_{S_x\sim D^m}\left[(\mathbf{x}_i)_j^2\right]} \text{ (linearity of the expectation)}
\end{aligned}$$

Based on the given D, we know $(\mathbf{x}_i)_j \sim \mathcal{N}(0,1)$. We know that

$$\begin{aligned}
\mathrm{Var}[(\mathbf{x}_i)_j] &\leq \mathbb{E}[(\mathbf{x}_i)_j^2] - \mathbb{E}[(\mathbf{x}_i)_j]^2 \\
1 &\leq \mathbb{E}[(\mathbf{x}_i)_j^2] - 0^2 \\
\implies \mathbb{E}[(\mathbf{x}_i)_j^2] &= 1
\end{aligned}$$

So:

$$\mathcal{R}_m(\mathcal{H}) \leq \frac{1}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{E}_{S_x\sim D^m}\left[(\mathbf{x}_i)_j^2\right]}$$

$$
\begin{aligned}
&= \frac{1}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} 1} \\
&= \frac{1}{m}\sqrt{mn} \\
\implies \mathcal{R}_m(\mathcal{H}) &\leq \sqrt{\frac{n}{m}}
\end{aligned}
$$

QED