Mehryar Mohri
Foundations of Machine Learning 2022
Courant Institute of Mathematical Sciences
Homework assignment 1
February 8, 2022
Due: February 22, 2022

A Concentration bound

1. We denote by $\mathcal{X}$ the input space and S an i.i.d sample of size m.

   (a) Show that there does not exist any hypothesis $h\colon \mathcal{X} \to \{0,1\}$ such that the following inequality holds with probability at least $e^{-m/3}$:

   $$R(h) - \widehat{R}_S(h) \geq \frac{1}{2}.$$

   (b) Suppose that the target concept to learn is $c \equiv 1$ and the target distribution D is the uniform distribution over the interval $[0,1]$. Design an algorithm such that for any sample S, the returned hypothesis $h_S\colon \mathcal{X} \to \{0,1\}$ satisfies the following equality:

   $$R(h_S) - \widehat{R}_S(h_S) = 1.$$

   (c) Why does part (b) not contradict part (a)?

Solution:

(a) By Hoeffding's inequality, for any hypothesis $h\colon \mathcal{X} \to \{0,1\}$, the following inequality holds:

$$\mathbb{P}\left[R(h) - \widehat{R}_S(h) \geq \frac{1}{2}\right] \leq e^{-m/2} < e^{-m/3}.$$

(b) The algorithm returns the hypothesis h_S defined by

$$h_S(x) = 1_{x \in S}.$$

Therefore, we have

$$\begin{aligned}\widehat{R}_S(h_S) &= \frac{1}{m}\sum_{i=1}^m 1_{h_S(x_i)=0} \\ &= \frac{1}{m}\sum_{i=1}^m 1_{x_i \notin S} \\ &= \frac{1}{m}\sum_{i=1}^m 0 \\ &= 0,\end{aligned}$$

and

$$\begin{aligned}R(h_S) &= \mathbb{P}_{x\sim D}[h_S(x) = 0] \\ &= \mathbb{P}_{x\sim D}[x \notin S] \\ &= 1.\end{aligned}$$

(c) Because h_S is not a fixed hypothesis. It depends on the sample S.

B PAC-Bayesian bound

1. Let $\mathcal{H}$ be a hypothesis set of functions mapping $\mathcal{X}$ to $\mathbb{R}$ and let ℓ be a loss function mapping $\mathbb{R} \times \mathcal{Y}$ to $[0,1]$. Denote the loss of a hypothesis h at point $z = (x,y) \in \mathcal{X} \times \mathcal{Y} = \mathcal{Z}$ by $L(h,z) = \ell(h(x),y)$. Let P and Q be probability measures over $\mathcal{H}$. In the PAC-Bayes framework, P represents the *prior probability* over the hypothesis class, i.e., the probability that a particular hypothesis is selected by the learning algorithm. Q represents the *posterior probability* selected after observing the training sample. In this exercise, we will derive learning bounds for randomized algorithms, in terms of the relative entropy of Q and P, denoted by $\mathsf{D}(Q \parallel P)$ (See E.2 of the textbook for the definition).

   (a) Define $\mathcal{G}_\mu$ via $\mathcal{G}_\mu = \{Q \in \Delta(\mathcal{H}) : \mathsf{D}(Q \parallel P) \leq \mu\}$, where we denote by $\Delta(\mathcal{H})$ the family of distributions over $\mathcal{H}$. Use the Rademacher complexity bound to show that for any $\delta > 0$, with probability at least $1 - \delta$, the following inequality holds for all $Q \in \mathcal{G}_\mu$:

   $$\mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^{m} L(h,z_i)\right] + 2\mathfrak{R}_m(\mathcal{G}_\mu) + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

   (b) It can be shown that the following inequality holds:

   $$\mathfrak{R}_m(\mathcal{G}_\mu) \leq \sqrt{\frac{2\mu}{m}}.$$

   Use this information to show that for any $\delta > 0$, with probability at least $1 - \delta$, the following inequality holds for all $Q \in \Delta(\mathcal{H})$:

   $$\mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^{m} L(h,z_i)\right] + \left(4 + \frac{1}{\sqrt{e}}\right)\sqrt{\frac{\max\{\mathsf{D}(Q \parallel P), 1\}}{m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

   (*Hint*: use the doubling trick, i.e., for some $a > 0$, $\Delta(\mathcal{H})$ can be written as the union of $\{Q \in \Delta(\mathcal{H}) : \mathsf{D}(Q \parallel P) \leq a\}$ and $\bigcup_{j=1}^{\infty}\left\{Q \in \Delta(\mathcal{H}) : a2^{j-1} < \mathsf{D}(Q \parallel P) \leq a2^j\right\}$. Then, use the union bound to extend the result in part (a). Note that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and $\frac{\log(2t)}{2} \leq \frac{t}{e}$ for $t > 0$.)

Solution:

(a) Note that the function $\mathbb{E}_{h\sim Q}[L(h,\cdot)]$ maps from $\mathcal{Z}$ to $[0,1]$. Then, by the Rademacher complexity bound, for any $\delta > 0$, with probability at least $1 - \delta$, the following inequality holds for all $Q \in \mathcal{G}_\mu$:

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}}\left[\mathop{\mathbb{E}}_{h\sim Q}[L(h,z)]\right] \leq \frac{1}{m}\sum_{i=1}^{m}\mathop{\mathbb{E}}_{h\sim Q}[L(h,z_i)] + 2\mathfrak{R}_m(\mathcal{G}_\mu) + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

(b) Part (a) along with the upper bound on $\mathfrak{R}_m(\mathcal{G}_\mu)$ imply that for any $\delta > 0$, with probability at least $1 - \delta$, the following inequality holds for all Q such that $\mathsf{D}(Q \parallel P) \leq \mu$:

$$\mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^{m} L(h,z_i)\right] + 2\sqrt{\frac{2\mu}{m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

For $j \geq 0$, define $\delta_j = 2^{-(j+1)}\delta$. Let $\Gamma_0 = \{Q \in \Delta(\mathcal{H}) : \mathsf{D}(Q \parallel P) \leq a\}$. For $j \geq 1$, let $\Gamma_j = \left\{Q \in \Delta(\mathcal{H}) : a2^{j-1} < \mathsf{D}(Q \parallel P) \leq a2^j\right\}$.

Therefore, by the union bound,

$$\begin{aligned}
&\mathbb{P}\left[\forall j \geq 0,\ \forall Q \in \Gamma_j,\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + 2\sqrt{\frac{2a2^j}{m}} + \sqrt{\frac{\log\frac{1}{\delta_j}}{2m}}\right]\\
&= 1 - \mathbb{P}\left[\exists j \geq 0,\ \exists Q \in \Gamma_j,\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] > \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + 2\sqrt{\frac{2a2^j}{m}} + \sqrt{\frac{\log\frac{1}{\delta_j}}{2m}}\right]\\
&\geq 1 - \sum_{j=0}^{\infty}\mathbb{P}\left[\exists Q \in \Gamma_j,\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] > \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + 2\sqrt{\frac{2a2^j}{m}} + \sqrt{\frac{\log\frac{1}{\delta_j}}{2m}}\right]\\
&= 1 - \sum_{j=0}^{\infty}\left(1 - \mathbb{P}\left[\forall Q \in \Gamma_j,\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + 2\sqrt{\frac{2a2^j}{m}} + \sqrt{\frac{\log\frac{1}{\delta_j}}{2m}}\right]\right)\\
&\geq 1 - \sum_{j=0}^{\infty}\delta_j\\
&= 1 - \delta.
\end{aligned}$$

For $j \geq 1$, if $Q \in \Gamma_j$, then $a2^j < 2\mathsf{D}(Q \parallel P)$ and $\delta_j \geq \frac{a\delta}{4\mathsf{D}(Q\parallel P)}$. Hence, for $j \geq 0$, if $Q \in \Gamma_j$, then

$$\begin{aligned}
&2\sqrt{\frac{2a2^j}{m}} + \sqrt{\frac{\log\frac{1}{\delta_j}}{2m}}\\
&\leq 4\sqrt{\frac{\max\{\mathsf{D}(Q\parallel P), a/2\}}{m}} + \sqrt{\frac{\log\max\{4\mathsf{D}(Q\parallel P)/a, 2\}}{2m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}} && \left(\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}\right)\\
&\leq 4\sqrt{\frac{\max\{\mathsf{D}(Q\parallel P), 1\}}{m}} + \sqrt{\frac{\log(2\max\{\mathsf{D}(Q\parallel P), 1\})}{2m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}} && (\text{take } a = 2)\\
&\leq \left(4 + \frac{1}{\sqrt{e}}\right)\sqrt{\frac{\max\{\mathsf{D}(Q\parallel P), 1\}}{m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}. && \left(\frac{\log(2t)}{2} \leq \frac{t}{e}\right)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
1-\delta &\leq \mathbb{P}\left[\forall j \geq 0,\ \forall Q \in \Gamma_j,\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + 2\sqrt{\frac{2a2^j}{m}} + \sqrt{\frac{\log\frac{1}{\delta_j}}{2m}}\right]\\
&\leq \mathbb{P}\left[\forall j \geq 0,\ \forall Q \in \Gamma_j,\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + \left(4 + \frac{1}{\sqrt{e}}\right)\sqrt{\frac{\max\{\mathsf{D}(Q\parallel P), 1\}}{m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}\right]\\
&= \mathbb{P}\left[\forall Q \in \Delta(\mathcal{H}),\ \mathop{\mathbb{E}}_{\substack{h\sim Q\\ z\sim\mathcal{D}}}[L(h,z)] \leq \mathop{\mathbb{E}}_{h\sim Q}\left[\frac{1}{m}\sum_{i=1}^m L(h,z_i)\right] + \left(4 + \frac{1}{\sqrt{e}}\right)\sqrt{\frac{\max\{\mathsf{D}(Q\parallel P), 1\}}{m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}\right].
\end{aligned}$$

C Rademacher complexity

1. Let $\mathcal{X} \subset \mathbb{R}^N$ and let $S = ((x_1, y_1), \ldots, (x_m, y_m)) \in (\mathcal{X} \times \mathcal{Y})^m$ be a sample of size m. In this problem, we consider the following linear hypothesis set

$$\mathcal{H} = \{x \mapsto w \cdot x : \|w\|_1 \leq \Lambda\}.$$

We denote by X the matrix $X = [x_1, \ldots, x_m]$ whose columns are the sample points. The (p,q)-group norm of a matrix M is defined as the q norm of the p norm of the columns of M, that is $\|M\|_{p,q} =$

$\left\| \left(\|M_1\|_p, \ldots, \|M_N\|_p \right) \right\|_q$, where M_is are the columns of M. We denote by $\{\sigma_i\}_{i=1}^m$ the Rademacher variables, that is independent uniform random variables taking values in $\{-1,+1\}$.

(a) Show that the empirical Rademacher complexity of $\mathcal{H}$ admits the following upper bound:

$$\widehat{\mathfrak{R}}_S(\mathcal{H}) \le \frac{\Lambda}{m}\sqrt{2\log(2N)}\left\|X^\top\right\|_{2,\infty}.$$

(*Hint*: use Massart's lemma.)

(b) Show that for any $0 < p < \infty$, there exists a positive constant C_p such that the following inequality holds for all $m \ge 1$ and real numbers $a_1, \ldots, a_m$.

$$\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right] \le C_p\left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}}$$

(*Hint*: For $p \le 2$, you can use Jensen's inequality. For $p > 2$, w.l.o.g., rescale such that $\sum_{i=1}^m a_i^2 = 1$, use the identity $\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}[X > t]dt$ for $X \ge 0$.)

(c) Show that for any $0 < p < \infty$, there exists a positive constant c_p such that the following inequality holds for all $m \ge 1$ and real numbers $a_1, \ldots, a_m$.

$$c_p\left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}} \le \mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right]$$

(*Hint*: For $p \ge 2$, you can use Jensen's inequality. For $p < 2$, use Hölder's inequality and part (b).)

(d) Use the inequality shown in part (c), show that the empirical Rademacher complexity of $\mathcal{H}$ admits the following lower bound:

$$\widehat{\mathfrak{R}}_S(\mathcal{H}) \ge c_1\frac{\Lambda}{m}\left\|X^\top\right\|_{2,\infty},$$

where c_1 is some positive constant in part (c) for $p = 1$.

(e) By providing an example, show that the dimension dependence of $\sqrt{\log N}$ in the upper bound in part (a) is tight (*Hint*: consider a data set with $N = 2^m$).

Solution:

(a) For any $i \in [m]$, we denote by x_{ij} the jth component of x_i.

$$\begin{aligned}
\widehat{\mathfrak{R}}_S(\mathcal{H}) &= \frac{1}{m}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|w\|_1 \le \Lambda}\sum_{i=1}^m \sigma_i w \cdot x_i\right] \\
&= \frac{1}{m}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|w\|_1 \le \Lambda} w\sum_{i=1}^m \sigma_i x_i\right] \\
&= \frac{\Lambda}{m}\mathop{\mathbb{E}}_{\sigma}\left[\left\|\sum_{i=1}^m \sigma_i x_i\right\|_\infty\right] && \text{(by def. of the dual norm)} \\
&= \frac{\Lambda}{m}\mathop{\mathbb{E}}_{\sigma}\left[\max_{j\in[N]}\left|\sum_{i=1}^m \sigma_i x_{ij}\right|\right] && \text{(by def. of } \|\cdot\|_\infty\text{)} \\
&= \frac{\Lambda}{m}\mathop{\mathbb{E}}_{\sigma}\left[\max_{j\in[N]}\max_{s\in\{-1,+1\}} s\sum_{i=1}^m \sigma_i x_{ij}\right] && \text{(by def. of abs. value)} \\
&= \frac{\Lambda}{m}\mathop{\mathbb{E}}_{\sigma}\left[\sup_{z\in A}\sum_{i=1}^m \sigma_i z_i\right],
\end{aligned}$$

where A denotes the set of vectors $\{s(x_{1j},\ldots,x_{mj})^\top : j \in [N], s \in \{-1,+1\}\}$. For any $z \in A$, we have $\sup_{z\in A}\|z\|_2 = \|X^\top\|_{2,\infty}$. Thus, by Massart's lemma, since A contains at most $2N$ elements, the following inequality holds:

$$\widehat{\mathfrak{R}}_S(\mathcal{H}) \le \Lambda \left\|X^\top\right\|_{2,\infty} \frac{\sqrt{2\log(2N)}}{m},$$

which concludes the proof.

(b) For $p \le 2$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right] &\le \left(\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^2\right]\right)^{\frac{p}{2}} && \text{(Jensen's inequality)}\\
&= \left(\mathop{\mathbb{E}}_{\sigma}\left[\sum_{i,j=1}^m \sigma_i\sigma_j(a_i a_j)\right]\right)^{\frac{p}{2}}\\
&= \left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}} && (\mathbb{E}[\sigma_i\sigma_j] = \mathbb{E}[\sigma_i]\,\mathbb{E}[\sigma_j] = 0 \text{ for } i \ne j)\\
&= C_p\left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}},
\end{aligned}$$

where $C_p = 1$. Next we consider the case where $p > 2$. Without loss of generality, rescale such that $\sum_{i=1}^m a_i^2 = 1$. Use the identity in the hint, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right] &= \int_0^{+\infty} \mathbb{P}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p > t\right]dt && \left(\mathbb{E}[|X|] = \int_0^{+\infty}\mathbb{P}[|X| > t]dt\right)\\
&= \int_0^{+\infty} \mathbb{P}\left[\left|\sum_{i=1}^m \sigma_i a_i\right| > t^{\frac{1}{p}}\right]dt\\
&\le 2\int_0^{+\infty} e^{-\frac{t^{\frac{2}{p}}}{2}}\,dt && \left(\sum_{i=1}^m a_i^2 = 1, \text{ Hoeffding's inequality}\right)\\
&= C_p\left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}},
\end{aligned}$$

where $C_p = 2\int_0^{+\infty} e^{-\frac{t^{\frac{2}{p}}}{2}}\,dt$.

(c) For $p \ge 2$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right] &\ge \left(\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^2\right]\right)^{\frac{p}{2}} && \text{(Jensen's inequality)}\\
&= \left(\mathop{\mathbb{E}}_{\sigma}\left[\sum_{i,j=1}^m \sigma_i\sigma_j(a_i a_j)\right]\right)^{\frac{p}{2}}\\
&= \left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}} && (\mathbb{E}[\sigma_i\sigma_j] = \mathbb{E}[\sigma_i]\,\mathbb{E}[\sigma_j] = 0 \text{ for } i \ne j)\\
&= c_p\left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}}
\end{aligned}$$

where $c_p = 1$. Next we consider the case where $p < 2$. Use the inequality shown in (b), we have

$$\begin{aligned}
\sum_{i=1}^m a_i^2 &= \mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^2\right] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^{\frac{2p}{3}} \left|\sum_{i=1}^m \sigma_i a_i\right|^{2-\frac{2p}{3}}\right] \\
&\leq \left(\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right]\right)^{\frac{2}{3}} \left(\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^{6-2p}\right]\right)^{\frac{1}{3}} && \text{(Hölder's inequality)} \\
&\leq \left(\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right]\right)^{\frac{2}{3}} C_{6-2p}^{\frac{1}{3}} \left(\sum_{i=1}^m a_i^2\right)^{1-\frac{p}{3}}. && \text{(by the ineq. shown in (b))}
\end{aligned}$$

Rearranging the terms, we obtain

$$\left(\frac{1}{C_{6-2p}}\right)^{\frac{1}{2}} \left(\sum_{i=1}^m a_i^2\right)^{\frac{p}{2}} \leq \mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i a_i\right|^p\right],$$

which concludes the proof.

(d) For any vector u, we denote by $|u|$ the vector derived from u by taking the absolute value of each of its components.

$$\begin{aligned}
\widehat{\mathfrak{R}}_S(\mathcal{H}) &= \frac{1}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|w\|_1 \leq \Lambda} \sum_{i=1}^m \sigma_i w \cdot x_i\right] \\
&= \frac{1}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|w\|_1 \leq \Lambda} w \sum_{i=1}^m \sigma_i x_i\right] \\
&= \frac{\Lambda}{m} \mathop{\mathbb{E}}_{\sigma}\left[\left\|\sum_{i=1}^m \sigma_i x_i\right\|_\infty\right] && \text{(by def. of the dual norm)} \\
&\geq \frac{\Lambda}{m} \left\|\mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i x_i\right|\right]\right\|_\infty && \text{(by sub-additivity of norm)} \\
&= \frac{\Lambda}{m} \max_{j \in [N]} \mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^m \sigma_i x_{ij}\right|\right] && \text{(by def. of } \|\cdot\|_\infty\text{)} \\
&\geq c_1 \frac{\Lambda}{m} \max_{j \in [N]} \left(\sum_{i=1}^m x_{ij}^2\right)^{\frac{1}{2}} && \text{(by the ineq. shown in (c))} \\
&= c_1 \frac{\Lambda}{m} \|X^\top\|_{2,\infty}.
\end{aligned}$$

(e) Consider a data set with $N = 2^m$. Take $\{x_i\}_{i=1}^m$ so that the rows of X are the set $\{-1, +1\}^m$. Then,

$\|X^\top\|_{2,\infty} = \sqrt{m}$ and the empirical Rademacher complexity can be computed as follows.

$$\begin{aligned}
\widehat{\mathfrak{R}}_S(\mathcal{H}) &= \frac{1}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|w\|_1 \le \Lambda} \sum_{i=1}^{m} \sigma_i w \cdot x_i\right] \\
&= \frac{1}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|w\|_1 \le \Lambda} w \sum_{i=1}^{m} \sigma_i x_i\right] \\
&= \frac{\Lambda}{m} \mathop{\mathbb{E}}_{\sigma}\left[\left\|\sum_{i=1}^{m} \sigma_i x_i\right\|_\infty\right] && \text{(by def. of the dual norm)} \\
&= \frac{\Lambda}{m} \mathop{\mathbb{E}}_{\sigma}\left[\max_{j \in [N]}\left|\sum_{i=1}^{m} \sigma_i x_{ij}\right|\right] && \text{(by def. of } \|\cdot\|_\infty\text{)} \\
&= \frac{\Lambda}{m} \mathop{\mathbb{E}}_{\sigma}[m] \\
&= \frac{\Lambda}{m\sqrt{\log 2}} \sqrt{\log(N)} \left\|X^\top\right\|_{2,\infty}. && (N = 2^m,\ \left\|X^\top\right\|_{2,\infty} = \sqrt{m})
\end{aligned}$$

Therefore, the dimension dependence of $\sqrt{\log N}$ in the upper bound is tight.