Name: Honda Li   Class: CSE 546 SPRING 2021   HW3B

B.1: Intro to Sample Complexity

B.1.a

Let's exam the statement:

$$\mathbb{P}\left[\hat{R}_n(f) = 0\right] = \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{f(x_i) \neq y_i\} = 0\right] \tag{B.1.a}$$
$$= \prod_{i=1}^{n} 1 - \mathbb{P}\left[f(x_i) \neq y_i\right]$$

Let's use the statement in the hypothesis. The statement was $R(f) > \epsilon$, which describes the event that $\mathbb{E}\left[\mathbf{1}\{f(x) \neq Y\}\right]$. so then $1 - R(f)$ describes exepected value of the event that: $1 - \mathbb{P}\left(f(x_i) \neq y\right)$. And notice that $1 - R(f) < 1 - \epsilon < \exp(epsilon)$, so then we can simplify the above expression into:

$$\prod_{i=1}^{n} 1 - \mathbb{P}\left[f(x_i) \neq y_i\right] = (1 - R(f)) \leq (\exp(\epsilon))^n = \exp(n\epsilon) \tag{B.1.a.1}$$
$$\implies \mathbb{P}\left[\hat{R}_n(f) = 0\right] \leq \exp(n\epsilon)$$

B.1.b

The results from the previous involves the hypothesis that $R(f) \geq \epsilon$, the theoretical risk of the model is larger than ϵ, therefore, under a larger scope the more appropriate inequality to make should be:

$$\mathbb{P}\left[\hat{R}_n(f) = 0 \wedge R(f) \geq \epsilon\right] \leq \exp(-n\epsilon) \tag{B.1.b.1}$$

For the proof, let's start with the following statement:

$$\mathbb{P}\left[f(f) > \epsilon \wedge \hat{R}_n(f) = 0\right] = \mathbb{P}\left[\hat{R}_n(f) = 0 \middle| R(f) > \epsilon\right]\mathbb{P}\left[R(f) > \epsilon\right] \tag{B.1.b.2}$$
$$\implies \mathbb{P}\left[f(f) > \epsilon \wedge \hat{R}_n(f) = 0\right] \leq \mathbb{P}\left[\hat{R}_n(f) = 0 \middle| R(f) > \epsilon\right] \leq \exp(-n\epsilon)$$

Now, the statement $\exists f \in \mathcal{F} : R(f) > \epsilon \wedge \hat{R}_n(f) = 0$ implies that occurence of at least one f that make the event true, hence, it's the union of the probability of each individual f in $\mathcal{F}$ which can make it true.

$$\mathbb{P}\left[\exists f \in \mathcal{F} : R(f) > \epsilon \wedge \hat{R}_n(f) = 0\right] = \mathbb{P}\left[\bigcup_{f\in\mathcal{F}}\left\{R(f) > \epsilon \wedge \hat{R}_n(f) = 0\right\}\right] \tag{B.1.b.3}$$
$$\underset{\text{Union Bound}}{\leq} \sum_{f\in\mathcal{F}}\mathbb{P}\left[R(f) > \epsilon \wedge \hat{R}_n(f) = 0\right]$$
$$\leq |\mathcal{F}|\exp(-n\epsilon)$$

B.1.c

Let's assume that there exists instance where the bounds for what we derived on part (b) could be strict (I believe it should.), then we can re-arrange and get the expression that:

$$\begin{aligned} |\mathcal{F}|\exp(-\epsilon n) &\leq \delta \qquad \text{(B.1.c.1)} \\ \ln(|\mathcal{F}|) - \epsilon n &\leq \ln(\delta) \\ -\epsilon n &\leq \ln(\delta) - \ln(|\mathcal{F}|) \\ \epsilon &\leq \frac{\ln(\delta) - \ln(|\mathcal{F}|)}{-n} \\ \epsilon &\leq \frac{\ln(\frac{|\mathcal{F}|}{\delta})}{n} \end{aligned}$$

Therefore, the largest ϵ is $\frac{1}{n}\ln(|\mathcal{F}|/\delta)$

B.1.d

We are proving the probability of the events $A \implies B$, where A, B are some kind of probabilistic events, Using some math we have: $\mathbb{P}(A \implies B) = 1 - \mathbb{P}(A \wedge \neg B)$, and in our case A is $\hat{R}_n(\hat{f}) = 0$ and B is $R(\hat{f}) - R(f^*) \leq \frac{\ln(|\mathcal{F}|/\delta)}{n}$.
Let's get into the math and I will label the steps and explain them.

$$\begin{aligned} \mathbb{P}\left[\widehat{R}_n(\hat{f}) = 0 \implies R(\hat{f}) - R(f^*) \leq \frac{\ln(|\mathcal{F}|/\delta)}{n}\right] &\underset{(1)}{=} 1 - \mathbb{P}\left[\widehat{R}_n(\hat{f}) = 0 \wedge R(\hat{f}) - R(f^*) \geq \frac{\ln(|\mathcal{F}|/\delta)}{n}\right] \qquad \text{(B.1.d.1)} \\ &\underset{(2)}{\geq} 1 - \mathbb{P}\left[\widehat{R}_n(\hat{f}) = 0 \wedge R(\hat{f}) \geq \frac{\ln(|\mathcal{F}|/\delta)}{n}\right] \\ &\underset{(3)}{\geq} 1 - \mathbb{P}\left[\widehat{R}_n(\hat{f}) = 0 \wedge R(\hat{f}) \geq \epsilon\right] \\ &\underset{(4)}{\geq} 1 - \mathbb{P}\left[\exists \hat{f} \in \mathcal{F} : \widehat{R}_n(\hat{f}) = 0 \wedge R(\hat{f}) \geq \epsilon\right] \\ &\underset{B.1.b.3}{\geq} 1 - |\mathcal{F}|\exp(-n\epsilon) \\ &\geq 1 - \delta \end{aligned}$$

(1) True becase: $A \implies B = \neg(A \wedge \neg B)$.

(2) True because $R(f^*) > 0$, meaning that $R(\hat{f}) - R(f^*) \leq \frac{1}{n}\ln(|\mathcal{F}|/\delta)$ implies that $R(\hat{f}) \geq \frac{1}{n}\ln(|\mathcal{F}|/\delta)$, which implies that the latter is a subset of the former, Therefore its probability is going to be higher. Justifying the inequality.

(3) True because $\epsilon < \frac{1}{n}\ln(|\mathcal{F}|/\delta)$, and this is just what we proved in (c) of B.1. And because of this, whatever $\hat{f}$ that makes $\hat{f} \geq \frac{1}{n}\ln(|\mathcal{F}|/\delta)$ will make $R(\hat{f}) \geq \epsilon$ as well, therefore the former is a subset of the latter, therefore the latter has a higher probability, justifying the inequality.

(4) The probability of the existence of at least one is higher than the probaility of the existence of any particular one, because the existence of any particular one has a prior that we chose that particular one to measure the imperical risk.

The last one just by the definition of δ.