Homework #3

CSE 446/546: Machine Learning
Ray Chen
Due: **Wednesday** Nov 23, 2022 11:59pm
A: 101 points, **B**: 16 points

B1:

- **Part a:**

$$
\begin{aligned}
d &= \parallel proj(x_0 - x) \parallel \\
&= \parallel \frac{(x_0 - X)w}{w \cdot w} \parallel \\
&= \parallel x_0 w - Xw \parallel \parallel \frac{1}{w} \parallel^2 \\
&= \frac{|x_0 w - Xw|}{\parallel w \parallel^2}
\end{aligned}
$$

Recall $w^T X + b = 0$, we have

$$d = \frac{|x_0^T w + b|}{\parallel w \parallel^2}$$

- **Part b:**

Suppose we have two hyperplanes, and we choose two point from each planes called $P1$ and $P2$.

By moving the two points with the distance of w. We have:

$$P1 + wt = P2 \tag{1}$$

Then we can find the $d = ||P1 - P2||$

$$
\begin{aligned}
& w^T(P1 - P2) = 2 \\
& \Rightarrow - \parallel w \parallel^2 t = 2 \\
& \Rightarrow t = \frac{-2}{\parallel w \parallel^2} \\
& \Rightarrow d = |t| \parallel w \parallel \\
& = \frac{2}{\parallel w \parallel^2} \parallel w \parallel \\
& = \frac{2}{\parallel w \parallel}
\end{aligned}
$$

B2:

a. **Part a:**

$$\begin{aligned} R(f) &= \mathbb{E}[]XY[\mathbf{1}(f(X) \neq Y)] \\ &= \mathbb{P}(f(X) \neq Y) > \epsilon \end{aligned}$$

Then, We have

$$\mathbb{P}(f(X) = Y) \leq 1 - \epsilon$$

.

So, we can obtian below

$$\begin{aligned} \mathbb{P}(\hat{R}_n(f) = 0) &\leq (1-\epsilon)^n \\ &\leq e^{-n\epsilon} \end{aligned}$$

b. **Part b:**

$$\begin{aligned} Pr(\exists f \in \mathcal{F} \text{ s.t. } R(f) > \epsilon \text{ and } \widehat{R}_n(f) = 0) &= Pr(A_1 \cup A_2 \cup \cdots \cup A_k), A_k K \in 1, 2, 3... \\ &\leq \sum_{1 \leq i \leq k} Pr(A_i) \\ &\leq \sum_{1 \leq i \leq k} e^{-n\epsilon} \\ &= |\mathcal{F}| e^{-n\epsilon} \end{aligned}$$

c. **Part c:**

$$\begin{aligned} &|\mathcal{F}| e^{-\epsilon n} \leq \delta \\ &\Rightarrow e^{-\epsilon n} \leq \frac{\delta}{|\mathcal{F}|} \\ &\Rightarrow \epsilon \geq \frac{1}{n} \ln \frac{|\mathcal{F}|}{\delta} \end{aligned}$$

d. **Part d:**

$$\begin{aligned} \mathbb{P}(\hat{R}_n(f) = 0 \Rightarrow R(\widehat{f}) - R(f^*) \leq \frac{1}{n} \log \frac{|\mathcal{F}|}{\delta}) &= \mathbb{P}(\hat{R}_n(f) = 0 \text{ and } R(\widehat{f}) - R(f^*) > \frac{1}{n} \log \frac{|\mathcal{F}|}{\delta}) \\ &= \mathbb{P}(\hat{R}_n(f) = 0 \text{ and } R(f) - R(f^*) > \epsilon^*) \\ &\geq 1 - \mathbb{P}\left[\hat{R}_n(f) = 0 \text{ and } R(f) > \epsilon^*\right] \\ &\geq 1 - \mathbb{P}\left[\exists f \in \mathcal{F} : R(f) > \epsilon^* \text{ and } \hat{R}_n(f) = 0\right] \\ &\geq 1 - |\mathcal{F}| e^{-\epsilon^* n} \\ &\geq 1 - \delta \end{aligned}$$