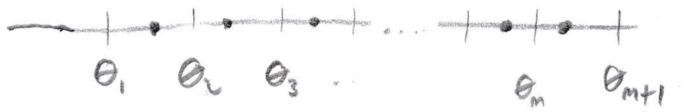
Math 597 Assignment 3

Jonathan Pearce
260672004

1) 

m points

given m points on the real line, there are m+1 options for threshold θ, m-1 options between 2 points, one option to the right of every point and one to the left of every point.

Given, one of these m+1 options for θ_i we can classify everything to the right as a positive instance (θ_i^R) or everything to the left as a positive instance (θ_i^L)

$\Longrightarrow$ $2(m+1)$ options for θ

observe that θ_1^R classifies every point the same way as θ_{m+1}^L. Similarily for θ_1^L and θ_{m+1}^R

$\Longrightarrow$ $2m$ options for θ

$$\boxed{\begin{aligned} &\Rightarrow \Pi_m(\mathcal{H}) \le 2m \\ &\Rightarrow \mathcal{R}_m(\mathcal{H}) \le \sqrt{\frac{2\log(2m)}{m}} \end{aligned}}$$

2) a) Suppose we are given an instance with one data point x_1. If x_1 is positivly labelled then h_{+1} can correctly classify it. If x_1 is negativly labelled then h_{-1} can correctly classify it.

Suppose we are given an instance with two data points x_1, x_2. Suppose x_1 is has a positive label and x_2 has a negative label. Neither h_{+1} or h_{-1} can correctly classify both points

$\Longrightarrow$ $\boxed{\text{VC - dimension} = 1}$

$$\mathcal{R}_S(\mathcal{H}) = \mathop{\mathbb{E}}_{\sigma}\left[\sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(x_i)\right]$$

$$= \mathop{\mathbb{E}}_{\sigma}\left[\sup_{h \in \mathcal{H}} \frac{1}{m} \left(\sum_{i=1}^{m} \sigma_i\right) h(x_1)\right]$$ Since $h(x_i) = h(x_j)$ $\forall i,j \in [h]$

$$= \mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m}\left|\sum_{i=1}^{m} \sigma_i\right|\right]$$

$$= \frac{1}{m} \mathop{\mathbb{E}}_{\sigma}\left[\left|\sum_{i=1}^{m} \sigma_i\right|\right]$$

$$\leq \frac{1}{m}\sqrt{\mathop{\mathbb{E}}_{\sigma}\left[\left(\sum_{i=1}^{m} \sigma_i\right)^2\right]}$$

$$= \frac{1}{m}\sqrt{\mathop{\mathbb{E}}_{\sigma}\left[\sum_{i=1}^{m} \sigma_i^2\right]}$$ $\mathbb{E}[\sigma_i \sigma_j] = 0 \quad i \neq j$

$$= \frac{1}{m}\sqrt{\sum_{i=1}^{m} 1}$$ $\mathbb{E}[\sigma_i^2] = 1$

$$\boxed{= \frac{1}{\sqrt{m}} = \sqrt{\frac{d}{m}} \quad \text{Since } d = 1}$$

b) Suppose we are given an instance with one data point y_1

If y_1 is labelled positive we can let $x_1 = y_1$ and then h_{+1} will correctly classify it. If y_1 is labelled negative then h_{-1} can correctly classify it

Suppose we are given an instance with two data points y_1, y_2 ($y_1 \neq y_2$). If y_1 and y_2 are both labelled positive no hypothesis in $\mathcal{H}$ can correctly classify both points

$$\boxed{\Longrightarrow \text{VC-dimension} = 1}$$

$$\mathfrak{R}_S(\mathcal{H}) = \mathbb{E}_\sigma\left[\sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(x_i)\right]$$

$$= \frac{1}{m} \mathbb{E}_\sigma\left[\sup_{h \in \mathcal{H}} \sigma_1 h(x_1) - \sum_{i=2}^{m} \sigma_i\right] \quad \text{Since } h(x_i) = -1 \text{ if } i \neq 1$$

$$= \frac{1}{m} \mathbb{E}_\sigma\left[\sup_{h \in \mathcal{H}} \sigma_1 h(x_1)\right] - \frac{1}{m} \mathbb{E}_\sigma\left[\sum_{i=2}^{m} \sigma_i\right]$$

$$= \frac{1}{m} \mathbb{E}_\sigma\left[\sup_{h \in \mathcal{H}} \sigma_1 h(x_1)\right] \quad \text{Since } \mathbb{E}[\sigma_i] = 0$$

$$= \frac{1}{m} \mathbb{E}_\sigma[1]$$

$$\boxed{= \frac{1}{m}}$$

3) a) $\mathcal{R}_m(\alpha\mathcal{H}) = \underset{S}{\mathbb{E}}\left[\underset{\sigma}{\mathbb{E}}\left[\sup_{h\in\alpha\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i)\right]\right]$

$$= \underset{S}{\mathbb{E}}\left[\underset{\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \alpha\sigma_i h(x_i)\right]\right]$$

Case $\alpha \geq 0$:

$$= \underset{S}{\mathbb{E}}\left[\underset{\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}} \frac{\alpha}{m}\sum_{i=1}^{m} \sigma_i h(x_i)\right]\right]$$

$$= \alpha\, \underset{S}{\mathbb{E}}\left[\underset{\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i)\right]\right]$$

$$= |\alpha|\, \mathcal{R}_m(\mathcal{H})$$

Case $\alpha < 0$:

$$= \underset{S}{\mathbb{E}}\left[\underset{\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}} \frac{(-\alpha)}{m}\sum_{i=1}^{m} (-\sigma_i) h(x_i)\right]\right]$$

note: σ_i and $-\sigma_i$ have the same distribution (this is stated on page 33 and 34 of the Foundations of Machine Learning by Mohri)

$$= |\alpha|\, \underset{S}{\mathbb{E}}\left[\underset{\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i)\right]\right]$$

$$= |\alpha|\, \mathcal{R}_m(\mathcal{H})$$

combining these two cases we get

$$\boxed{\Rightarrow \mathcal{R}_m(\alpha\mathcal{H}) = |\alpha|\, \mathcal{R}_m(\mathcal{H})}$$

b)

$$\mathfrak{R}_m(\mathcal{H}+\mathcal{H}') = \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H},\, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^{m} \sigma_i\left(h(x_i) + h(x_i')\right)\right]\right]$$

$$= \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H},\, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i) + \sup_{h\in\mathcal{H},\, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h'(x_i)\right]\right]$$

$$= \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i)\right]\right] + \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h'(x_i)\right]\right]$$

$$= \mathfrak{R}_m(\mathcal{H}) + \mathfrak{R}_m(\mathcal{H}')$$

c) $\mathfrak{R}_m(\max(h,h')) = \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m \sigma_i \max(h(x_i), h'(x_i))\right]\right]$

Use $\max(a,b) = \frac{1}{2}\left[a + b + |a-b|\right]$

$$= \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}, h'\in\mathcal{H}} \frac{1}{2m}\sum_{i=1}^m \sigma_i \left(h(x_i) + h'(x_i) + |h(x_i) - h'(x_i)|\right)\right]\right]$$

$$= \frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i)\right]\right] + \frac{1}{2}\left[\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m \sigma_i h'(x_i)\right]\right]\right]$$

$$+ \frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m \sigma_i |h(x_i) - h'(x_i)|\right]\right]$$

$$= \frac{1}{2}\mathfrak{R}_m(\mathcal{H}) + \frac{1}{2}\mathfrak{R}_m(\mathcal{H}') + \underbrace{\frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m \sigma_i |h(x_i) - h'(x_i)|\right]\right]}_{\text{denote } ①}$$

<u>Bounding ①</u>

let $f(x) = |x|$, $|f(x) - f(y)| = ||x| - |y|| \leq |x-y|$ (by the reverse Δ - inequality)

$\boxed{\Rightarrow \text{absolute value function is 1-lipschitz}}$

$\Rightarrow$ using Talgrand's Lemma

$$① \leq \frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m \sigma_i (h(x_i) - h'(x_i))\right]\right]$$

$$= \frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}, h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i) + \sup_{h\in\mathcal{H}, h'\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^m -\sigma_i h'(x_i)\right]\right]$$

$$= \frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i)\right]\right] + \frac{1}{2}\mathbb{E}_S\left[\mathbb{E}_\sigma\left[\sup_{h'\in\mathcal{H}'} \frac{1}{m}\sum_{i=1}^m -\sigma_i h'(x_i)\right]\right]$$

Use the fact that σ_i and $-\sigma_i$ follow the same distribution (Mohri page 33/34)

$$= \frac{1}{2} \mathop{E}_{S}\left[\mathop{E}_{\sigma}\left[\sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(x_i)\right]\right] + \frac{1}{2} \mathop{E}_{S}\left[\mathop{E}_{\sigma}\left[\sup_{h' \in \mathcal{H}'} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h'(x_i)\right]\right]$$

$$= \frac{1}{2} \mathfrak{R}_m(\mathcal{H}) + \frac{1}{2} \mathfrak{R}_m(\mathcal{H}')$$

$$\boxed{\begin{aligned} \Rightarrow \mathfrak{R}_m(\max(h,h') : h \in \mathcal{H}, h' \in \mathcal{H}') &\leq \frac{1}{2}\mathfrak{R}_m(\mathcal{H}) + \frac{1}{2}\mathfrak{R}_m(\mathcal{H}') + \frac{1}{2}\mathfrak{R}_m(\mathcal{H}) + \frac{1}{2}\mathfrak{R}_m(\mathcal{H}') \\ &= \mathfrak{R}_m(\mathcal{H}) + \mathfrak{R}_m(\mathcal{H}') \end{aligned}}$$

4) a) $$\hat{\mathfrak{R}}_s(\mathcal{H}) = \mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m} \sup_{\|w\|_1 \leq \Lambda',\ \|u_j\|_2 \leq \Lambda} \sum_{i=1}^{m} \sigma_i \sum_{j=1}^{n_2} w_j \sigma(u_j \cdot x_i)\right]$$

$$= \mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m} \sup_{\|w\|_1 \leq \Lambda';\ \|u_j\|_2 \leq \Lambda} \sum_{j=1}^{n_2} w_j \sum_{i=1}^{m} \sigma_i \sigma(u_j \cdot x_i)\right]$$

$$= \mathop{\mathbb{E}}_{\sigma}\left[\frac{\Lambda'}{m} \sup_{\|u_j\| \leq \Lambda} \max_{j \in [1, n_2]} \left|\sum_{i=1}^{m} \sigma_i \sigma(u_j \cdot x_i)\right|\right]$$

$$= \frac{\Lambda'}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|u_j\| \leq \Lambda,\ j \in [1, n_2]} \left|\sum_{i=1}^{m} \sigma_i \sigma(u_j \cdot x_i)\right|\right]$$

$$= \frac{\Lambda'}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|u_j\| \leq \Lambda} \left|\sum_{i=1}^{m} \sigma_i \sigma(u_j \cdot x_i)\right|\right]$$

b) σ is L-lipSchitz, by Talagrand's lemma

$$\hat{\mathfrak{R}}_s(\mathcal{H}) \leq \frac{\Lambda' L}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|u_j\| \leq \Lambda} \left|\sum_{i=1}^{m} \sigma_i (u \cdot x_i)\right|\right]$$

$$= \frac{\Lambda' L}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|u_j\| \leq \Lambda} \sup_{s \in \{-1, 1\}} s \sum_{i=1}^{m} \sigma_i (u \cdot x_i)\right]$$

$$= \frac{\Lambda' L}{m} \mathop{\mathbb{E}}_{\sigma}\left[\sup_{\|u_j\| \leq \Lambda\ s \in \{-1, 1\}} s \sum_{i=1}^{m} \sigma_i (u \cdot x_i)\right]$$

$$= \Lambda' L \hat{\mathfrak{R}}_s(\mathcal{H}')$$

c) $\hat{\mathfrak{R}}_S(H') = \frac{1}{m} \mathbb{E}_{\sigma}\left[\sup_{\|u\|_2 \leq \Lambda,\, s \in \{-1,1\}} s \sum_{i=1}^{m} \sigma_i (u \cdot x_i)\right]$

$$= \frac{1}{m} \mathbb{E}_{\sigma}\left[\sup_{\|u\|_2 \leq \Lambda} \left|\sum_{i=1}^{m} \sigma_i (u \cdot x_i)\right|\right]$$

$$= \frac{1}{m} \mathbb{E}_{\sigma}\left[\sup_{\|u\|_2 \leq \Lambda} \left|u \cdot \sum \sigma_i x_i\right|\right]$$

let $u = \dfrac{\Lambda \sum_{i=1}^{m} \sigma_i x_i}{\|\sum_{i=1}^{m} \sigma_i x_i\|_2}$

using Cauchy-Schwarz,

$$= \frac{\Lambda}{m} \mathbb{E}_{\sigma}\left[\left\|\sum_{i=1}^{m} \sigma_i x_i\right\|_2\right]$$

d) $\hat{\mathfrak{R}}_S(\mathcal{H}') = \frac{\Lambda}{m} \mathop{\mathbb{E}}_{\sigma}\left[\|\sum_{i=1}^{m} \sigma_i x_i\|_2\right]$

$$\leq \frac{\Lambda}{m}\sqrt{\mathop{\mathbb{E}}_{\sigma}\left[\|\sum_{i=1}^{m} \sigma_i x_i\|_2^2\right]}$$

$$= \frac{\Lambda}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m} \mathop{\mathbb{E}}_{\sigma}[\sigma_i \sigma_j](x_i \cdot x_j)}$$

$$= \frac{\Lambda}{m}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m} \mathbb{1}_{\{i=j\}}(x_i \cdot x_j)}$$

$$= \frac{\Lambda}{m}\sqrt{\sum_{i=1}^{m} \|x_i\|_2^2}$$

e) $\hat{\mathfrak{R}}_{T}(\mathcal{H}) \leq \Lambda' L \hat{\mathfrak{R}}_S(\mathcal{H}')$

$$\leq \frac{\Lambda' \Lambda L}{m}\sqrt{\sum_{i=1}^{m} \|x_i\|_2^2}$$

$$\leq \frac{\Lambda' \Lambda L}{m}\sqrt{\sum_{i=1}^{m} r^2}$$

$$= \frac{\Lambda' \Lambda L}{m}\sqrt{m r^2}$$

$$= \frac{\Lambda' \Lambda L r}{\sqrt{m}}$$

5) Suppose we have an instance with $2k$ points

Claim: There are at most k groups of consecutive positive points within the $2k$ points. Where a group of consecutive positive points is one or more points that are all classified positivly and have no negativly classified points between them

Proof: Assume there are $k+1$ groups of consecutive positive points within the $2k$ points

$\Rightarrow$ there are at least k negative points, Since each group of positive points is seperated by at least one negative point

Since each of the $k+1$ groups of consecutive positive points must have at least one positive point

$\Rightarrow$ there are at least $k+1$ positive points

$\Rightarrow$ there are at least $2k+1$ points, which is a contradiction

$\Rightarrow$ Each of the (at most) k groups of consecutive positive points can be covered by one of the k intervals

$\Rightarrow$ Union of k intervals shatters $2k$ points

Suppose we have an instance with $2k+1$ points, $x_1, x_2, \ldots x_{2k}, x_{2k+1}$ with $x_1 < x_2 < x_3 < \ldots < x_{2n} < x_{2k+1}$

Suppose we label the points as follows

$$x_i = \begin{cases} 1 & i \text{ odd} \\ -1 & i \text{ even} \end{cases} \Rightarrow + - + - + - \ldots - + - +$$

$\Rightarrow$ $k+1$ positive instances, with each consecutive instance seperated by a negative instance

$\Rightarrow$ $k+1$ groups of consecutive positive points

$\Rightarrow$ $k+1$ intervals are required to correctly classify all $2k+1$ points

$\Rightarrow$ union of k interval <u>does not</u> shatter $2k+1$ points

$$\boxed{\Rightarrow \text{VC-dimension} = 2k}$$

6) consider the instance $\{1/2, 5/4, 2\}$

<u>Classification</u>

$-$	$-$	$-$	$\alpha = 3$
$-$	$-$	$+$	$\alpha = \frac{3}{2}$
$-$	$+$	$-$	$\alpha = \frac{3}{4}$
$+$	$-$	$-$	$\alpha = \frac{1}{6}$
$-$	$+$	$+$	$\alpha = \frac{6}{5}$
$+$	$-$	$+$	$\alpha = \frac{-1}{4}$
$+$	$+$	$-$	$\alpha = \frac{1}{3}$
$+$	$+$	$+$	$\alpha = -2$

$\Rightarrow$ $\boxed{\{\frac{1}{2}, \frac{5}{4}, 2\} \text{ can be shattered}}$

now suppose we have an instance with 4 points x_1, x_2, x_3, x_4 with $x_1 < x_2 < x_3 < x_4$ and the classification $+ - + -$

clearly if $x_1 \in [\alpha+2, \infty)$ our classification will be wrong

$\Rightarrow x_1 \in [\alpha, \alpha+1]$ and is correctly classified $+$

$\Rightarrow x_2 \in (\alpha+1, \alpha+2)$ and is correctly classified $-$

$\Rightarrow x_3 \in [\alpha+2, +\infty)$ and is correctly classified $+$

$\Rightarrow x_4 \in [\alpha+2, +\infty)$ and is <u>incorrectly</u> classified $+$

$\Rightarrow$ no set of 4 points can be shattered

$\boxed{\Rightarrow \text{VC-dimension} = 3}$

7) $\mathcal{R}_m(\mathcal{H}^\varepsilon) = \underset{S,\sigma}{\mathbb{E}}\left[\sup_{f\in\mathcal{H}^\varepsilon} \frac{1}{m}\sum_{i=1}^{m} \sigma_i f(x_i)\right]$

Because in $\mathcal{H}^\varepsilon$, there exists $h \in \mathcal{H}^0$ such that $\mathbb{P}[f(x) \neq h(x)] \leq \varepsilon$ empirically we expect $f(x_i) \neq h(x_i)$ in less than $m \cdot \varepsilon$ instances of the data

$\Rightarrow$ we expect $f(x_i) = h(x_i)$ in at least $m(1-\varepsilon)$ instances of the data

To make it work with summation indexes we round $m(1-\varepsilon)$ to the nearest integer

$$\lceil m(1-\varepsilon)\rceil = \lceil m - m\varepsilon\rceil = m - \lceil m\varepsilon\rceil = m\left(1 - \frac{\lceil m\varepsilon\rceil}{m}\right) = m(1-\tilde{\varepsilon})$$

$$= \underset{S,\sigma}{\mathbb{E}}\left[\sup_{f\in\mathcal{H}^\varepsilon}\left(\frac{1}{m}\sum_{i=1}^{(1-\tilde{\varepsilon})m} \sigma_i f(x_i) + \frac{1}{m}\sum_{i=(1-\tilde{\varepsilon})m+1}^{m} \sigma_i f(x_i)\right)\right]$$

$$\leq \underset{S,\sigma}{\mathbb{E}}\left[\sup_{f\in\mathcal{H}^\varepsilon} \frac{1}{m}\sum_{i=1}^{(1-\tilde{\varepsilon})m} \sigma_i f(x_i) + \sup_{f\in\mathcal{H}^\varepsilon} \frac{1}{m}\sum_{i=(1-\tilde{\varepsilon})m+1}^{m} \sigma_i f(x_i)\right]$$

$$= \underset{S,\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}^0} \frac{1}{m}\sum_{i=1}^{(1-\tilde{\varepsilon})m} \sigma_i h(x_i) + \sup_{f\in\mathcal{H}^\varepsilon} \frac{1}{m}\sum_{i=(1-\tilde{\varepsilon})m+1}^{m} \sigma_i f(x_i)\right]$$

$$= \underset{S,\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}^0} \frac{1}{m}\sum_{i=1}^{(1-\tilde{\varepsilon})m} \sigma_i h(x_i)\right] + \underset{S,\sigma}{\mathbb{E}}\left[\sup_{f\in\mathcal{H}^\varepsilon} \frac{1}{m}\sum_{i=(1-\tilde{\varepsilon})m+1}^{m} \sigma_i f(x_i)\right]$$

$$\leq \underset{S,\sigma}{\mathbb{E}}\left[\sup_{h\in\mathcal{H}^0} \frac{1}{(1-\tilde{\varepsilon})m}\sum_{i=1}^{(1-\tilde{\varepsilon})m} \sigma_i h(x_i)\right] + \mathbb{E}\left[\frac{1}{m}\sum_{i=(1-\tilde{\varepsilon})m+1}^{m} 1\right] \quad \left(\text{Since } \sup \sigma_i f(x_i) = 1\right)$$

$$= \mathcal{R}_{(1-\tilde{\varepsilon})m}(\mathcal{H}^0) + \frac{1}{m}\left[m - m + \tilde{\varepsilon}m - 1\right]$$

$$\boxed{\leq \mathcal{R}_{(1-\tilde{\varepsilon})m}(\mathcal{H}^0) + \tilde{\varepsilon}}$$