Online Learning: Sequential Rademacher Complexity Bounded by Integrated Complexity (2)

Yuantong Li

Purdue University

li3551@purdue.edu

Nov 21, 2019

Online Learning Model

1. *Online Learning Model*:

   Let $\mathcal{F}$ be a class of functions and $\mathcal{X}$ some set. The Online Learning Model is defined as the following T-round interaction between the learner and the adversary: On round $t = 1, ..., T$, the learner chooses $f_t \in \mathcal{F}$, the adversary picks $x_t \in \mathcal{X}$, and the learner suffers loss $f_t(x_t)$. At the end of T rounds we define *regret*

   $$\mathbf{R}\left(f_{1:T}, x_{1:T}\right) = \sum_{t=1}^{T} f_t\left(x_t\right) - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} f\left(x_t\right)$$

   as the difference between the cumulative loss of the player as compared to the cumulative loss of the best fixed comparator.

2. *Online learnable*:

   For the given pair $(\mathcal{F}, \mathcal{X})$, the problem is said to be online learnable if there exists an algorithm for the learner such that regret grows sublinearly.

Value of the Game

Theorem 1

Let $\mathcal{F}$ and $\mathcal{X}$ be the sets of moves for the two players, satisfying the necessary conditions for the minimax theorem to hold. Denote by $\mathcal{Q}$ and $\mathcal{P}$ the sets of probability distributions (mixed strategies) on $\mathcal{F}$ and $\mathcal{X}$, respectively. Then

$$\begin{aligned}\mathcal{V}_T(\mathcal{F},\mathcal{X}) &= \inf_{q_1\in\mathcal{Q}}\sup_{x_1\in\mathcal{X}}\mathbb{E}_{f_1\sim q_1}\dots\inf_{q_T\in\mathcal{Q}}\sup_{x_T\in\mathcal{X}}\mathbb{E}_{f_T\sim q_T}[\sum_{t=1}^T f_t(x_t)-\inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)]\\ &= \sup_{p_1}\mathbb{E}_{x_1\sim p_1}\dots\sup_{p_T}\mathbb{E}_{x_T\sim p_T}[\sum_{t=1}^T\inf_{f_t\in\mathcal{F}}\mathbb{E}_{x_t\sim p_t}[f_t(x_t)]-\inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)]\end{aligned}$$

- $\mathcal{F}$: is a subset of a separable metric space.
- $\mathcal{Q}$: the set of probability distributions on $\mathcal{F}$.
- p_t: the distribution on x_t

Some definitions

Definition 1. Online Learnable (Formal)

A class $\mathcal{F}$ is said to be online learnable with respect to the given $\mathcal{X}$ if

$$\limsup_{T\to\infty} \frac{\mathcal{V}_T(\mathcal{F},\mathcal{X})}{T} = 0$$

Since complexity of $\mathcal{F}$ is the focus, we shall often write $\mathcal{V}(\mathcal{F})$, and the dependence on $\mathcal{X}$ will be implicit.

Definition 2. Sequential Rademacher Complexity (SRC)

The Sequential Rademacher Complexity of a function class $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ is defined as

$$\mathfrak{R}_T(\mathcal{F}) = \sup_{\mathbf{x}} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \epsilon_t f\left(\mathbf{x}_t(\epsilon)\right) \right]$$

where the outer supremum is taken over all $\mathcal{X}$-valued trees of depth T and $\epsilon = (\epsilon_1, \ldots, \epsilon_T)$ is a sequence of i.i.d. Rademacher random variables.

Upper bound by SRC

Theorem 2

The minimax value of a randomized game is bounded as

$$\mathcal{V}_T(\mathcal{F}) \leq 2\mathfrak{R}_T(\mathcal{F}) \tag{1}$$

Theorem 3

For any function class, $\mathcal{F} \subseteq [-1, 1]^{\mathcal{X}}$

$$\mathfrak{R}_T(\mathcal{F}) \leq \mathfrak{D}_T(\mathcal{F}) \tag{2}$$

Definition 3

The *Integrated Complexity* of a function class $\mathcal{F} \subseteq [-1, 1]^{\mathcal{X}}$ is defined as

$$\mathfrak{D}_T(\mathcal{F}) = \inf_{\alpha} \left\{ 4T\alpha + 12 \int_{\alpha}^{1} \sqrt{T \log \mathcal{N}_2(\delta, \mathcal{F}, T)}\, d\delta \right\} \tag{3}$$

Finite Class Lemma

Lemma 1

For any finite set V of $\mathbb{R}$-valued trees of depth T we have that

$$\mathbb{E}_\epsilon\left[\max_{\mathbf{v}\in V}\sum_{t=1}^{T}\epsilon_t\mathbf{v}_t(\epsilon)\right]\leq\sqrt{2\log(|V|)\max_{\mathbf{v}\in V}\max_{\epsilon\in\{\pm 1\}^T}\sum_{t=1}^{T}\mathbf{v}_t(\epsilon)^2}$$

A simple consequence of the above lemma is that if $\mathcal{F}\subseteq[-1,1]^{\mathcal{X}}$ is a finite class, then for any given tree $\mathbf{x}$ we have that

$$\mathbb{E}_\epsilon\left[\max_{f\in\mathcal{F}}\sum_{t=1}^{T}\epsilon_t f\left(\mathbf{x}_t(\epsilon)\right)\right]\leq\mathbb{E}_\epsilon\left[\max_{\mathbf{v}\in\mathcal{F}(\mathbf{x})}\sum_{t=1}^{T}\epsilon_t\mathbf{v}_t(\epsilon)\right]\leq\sqrt{2T\log(|\mathcal{F}|)}$$

Finite Class Lemma

Proof of Lemma (1)

For any $\lambda > 0$, we invoke Jensen's inequality to get

$$M(\lambda) := \exp\left\{\lambda \mathbb{E}_\epsilon \left[\max_{\mathbf{v}\in V} \sum_{t=1}^{T} \epsilon_t \mathbf{v}_t(\epsilon)\right]\right\} \leq \mathbb{E}_\epsilon \left[\exp\left\{\lambda \max_{\mathbf{v}\in V} \sum_{t=1}^{T} \epsilon_t \mathbf{v}_t(\epsilon)\right\}\right]$$

$$\leq \mathbb{E}_\epsilon \left[\max_{\mathbf{v}\in V} \exp\left\{\lambda \sum_{t=1}^{T} \epsilon_t \mathbf{v}_t(\epsilon)\right\}\right] \leq \mathbb{E}_\epsilon \left[\sum_{\mathbf{v}\in V} \exp\left\{\lambda \sum_{t=1}^{T} \epsilon_t \mathbf{v}_t(\epsilon)\right\}\right]$$

With the usual technique of peeling from the end,

$$M(\lambda) \leq \sum_{\mathbf{v}\in V} \mathbb{E}_{\epsilon_1,\ldots,\epsilon_T} \left[\prod_{t=1}^{T} \exp\left\{\lambda \epsilon_t \mathbf{v}_t\left(\epsilon_{1:t-1}\right)\right\}\right]$$

$$= \sum_{\mathbf{v}\in V} \mathbb{E}_{\epsilon_1,\ldots,\epsilon_{T-1}} \left[\prod_{t=1}^{T-1} \exp\left\{\lambda \epsilon_t \mathbf{v}_t\left(\epsilon_{1:t-1}\right)\right\} \times A\right]$$

where $A = \left(\frac{\exp\left\{\lambda \mathbf{v}_T\left(\epsilon_{1:T-1}\right)\right\} + \exp\left\{-\lambda \mathbf{v}_T\left(\epsilon_{1:T-1}\right)\right\}}{2}\right)$

Finite Class Lemma

Proof of Lemma (2)

$$\leq \sum_{\mathbf{v}\in V} \mathbb{E}_{\epsilon_1,\ldots,\epsilon_{T-1}} \left[\prod_{t=1}^{T-1} \exp\left\{\lambda \epsilon_t \mathbf{v}_t\left(\epsilon_{1:t-1}\right)\right\} \times \exp\left\{ \frac{\lambda^2 \mathbf{v}_T\left(\epsilon_{1:T-1}\right)^2}{2} \right\} \right]$$

where we used the inequality $\frac{1}{2}\{\exp(a) + \exp(-a)\} \leq \exp(a^2/2)$, valid for all $a \in \mathbb{R}$. Peeling off the second term is a bit more involved:

$$M(\lambda) \leq \sum_{\mathbf{v}\in V} \mathbb{E}_{\epsilon_1,\ldots,\epsilon_{T-2}} \left[\prod_{t=1}^{T-2} \exp\left\{\lambda \epsilon_t \mathbf{v}_t\left(\epsilon_{1:t-1}\right)\right\} \times \right.$$

$$\frac{1}{2}\left(\exp\left\{\lambda \mathbf{v}_{T-1}\left(\epsilon_{1:T-2}\right)\right\} \exp\left\{ \frac{\lambda^2 \mathbf{v}_T\left(\left(\epsilon_{1:T-2}, 1\right)\right)^2}{2} \right\} \right.$$

$$\left.\left. + \exp\left\{-\lambda \mathbf{v}_{T-1}\left(\epsilon_{1:T-2}\right)\right\} \exp\left\{ \frac{\lambda^2 \mathbf{v}_T\left(\left(\epsilon_{1:T-2}, -1\right)\right)^2}{2} \right\} \right) \right]$$

Finite Class Lemma

Proof of Lemma (3)

Consider the term inside:

$$\frac{1}{2}\left(\exp\left\{\lambda \mathbf{v}_{T-1}\left(\epsilon_{1:T-2}\right)\right\}\exp\left\{\frac{\lambda^2 \mathbf{v}_T\left(\left(\epsilon_{1:T-2},1\right)\right)^2}{2}\right\}\right)+$$

$$\frac{1}{2}\left(\exp\left\{-\lambda \mathbf{v}_{T-1}\left(\epsilon_{1:T-2}\right)\right\}\exp\left\{\frac{\lambda^2 \mathbf{v}_T\left(\left(\epsilon_{1:T-2},-1\right)\right)^2}{2}\right\}\right)$$

$$\leq \max_{\varepsilon_{T-1}}\left(\exp\left\{\frac{\lambda^2 \mathbf{v}_T\left(\left(\epsilon_{1:T-2},\epsilon_{T-1}\right)\right)^2}{2}\right\}\right)\times$$

$$\frac{\exp\left\{\lambda \mathbf{v}_{T-1}\left(\epsilon_{1;T-2}\right)\right\}+\exp\left\{-\lambda \mathbf{v}_{T-1}\left(\epsilon_{1;T-2}\right)\right\}}{2}$$

$$\leq \max_{\epsilon_{T-1}}\left(\exp\left\{\frac{\lambda^2 \mathbf{v}_T\left(\left(\epsilon_{1;T-2},\epsilon_{T-1}\right)\right)^2}{2}\right\}\right)\exp\left\{\frac{\lambda^2 \mathbf{v}_{T-1}\left(\epsilon_{1:T-2}\right)^2}{2}\right\}$$

$$=\exp\left\{\frac{\lambda^2 \max_{\epsilon_{T-1}\in\{\pm 1\}}\left(\mathbf{v}_{T-1}\left(\epsilon_{1:T-2}\right)^2+\mathbf{v}_T\left(\epsilon_{1:T-1}\right)^2\right)}{2}\right\}$$

Finite Class Lemma

Proof of Lemma (4)

Repeating the last steps, we show that for any i,

$$M(\lambda) \leq \sum_{\mathbf{v}\in V} E_{\epsilon_1,\ldots,\epsilon_{i-1}}[\prod_{t=1}^{i-1} \exp\left\{\lambda\epsilon_t \mathbf{v}_t\left(\epsilon_{1:t-1}\right)\right\} \times$$
$$\exp\left\{\frac{\lambda^2 \max_{\epsilon_i\ldots\epsilon_{T-1}\in\{\pm1\}} \sum_{t=i}^{T} \mathbf{v}_t\left(\epsilon_{1:t-1}\right)^2}{2}\right\}]$$

We arrive at

$$M(\lambda) \leq \sum_{\mathbf{v}\in V} \exp\left\{\frac{\lambda^2 \max_{\epsilon_1\ldots\epsilon_{T-1}\in\{\pm1\}} \sum_{t=1}^{T} \mathbf{v}_t\left(\epsilon_{1:t-1}\right)^2}{2}\right\}$$
$$\leq |V| \exp\left\{\frac{\lambda^2 \max_{\mathbf{v}\in V} \max_{\epsilon\in\{\pm1\}} \sum_{t=1}^{T} \mathbf{v}_t(\epsilon)^2}{2}\right\}$$

Finite Class Lemma

Proof of Lemma (5)

Take logarithms on both sides, dividing by λ and setting

$$\lambda = \sqrt{\frac{2\log(|V|)}{\max_{\mathbf{v}\in V}\max_{\epsilon\in\{\pm1\}} T\sum_{t=1}^{T}\mathbf{v}_t(\epsilon)^2}}$$

we conclude that

$$\mathbb{E}_{\epsilon_1,\dots,\epsilon_T}\left[\max_{\mathbf{v}\in V}\sum_{t=1}^{T}\epsilon_t\mathbf{v}_t(\epsilon)\right] \le \sqrt{2\log(|V|)\max_{\mathbf{v}\in V}\max_{\epsilon\in\{\pm1\}^T}\sum_{t=1}^{T}\mathbf{v}_t(\epsilon)^2}$$

Proof of the Theorem (1)

Define $\beta_0 = 1$ and $\beta_j = 2^{-j}$. For a fixed tree $\mathbf{x}$ of depth T, let V_j be an ℓ_2-cover at scale β_j. For any path $\epsilon \in \{\pm 1\}^T$ and any $f \in \mathcal{F}$, let $\mathbf{v}[f,\epsilon]^j \in V_j$ the element of the cover such that

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\mathbf{v}[f,\epsilon]_t^j(\epsilon) - f\left(\mathbf{x}_t(\epsilon)\right)\right|^2} \leq \beta_j$$

where $\mathbf{v}[f,\epsilon]_t^j$ denotes the t-th mapping of $\mathbf{v}[f,\epsilon]^j$. For any $t \in [T]$, we have

$$f\left(\mathbf{x}_t(\epsilon)\right) = f\left(\mathbf{x}_t(\epsilon)\right) - \mathbf{v}[f,\epsilon]_t^N(\epsilon) + \sum_{j=1}^{N}\left(\mathbf{v}[f,\epsilon]_t^j(\epsilon) - \mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon)\right)$$

where $\mathbf{v}[f,\epsilon]_t^0(\epsilon) = 0$. Hence,

$$E_\epsilon[\sup_{f\in\mathcal{F}}\sum_{t=1}^{T}\epsilon_t f(\mathbf{x}_t(\epsilon))] = E_\epsilon[\sup_{f\in\mathcal{F}}\sum_{t=1}^{T}\epsilon_t(f\left(\mathbf{x}_t(\epsilon)\right) - \mathbf{v}[f,\epsilon]_t^N(\epsilon)$$
$$+ \sum_{j=1}^{N}(\mathbf{v}[f,\epsilon]_t^j(\epsilon) - \mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon)))$$

Proof of the Theorem (2)

$$
\begin{aligned}
&= \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \epsilon_t \left(f(\mathbf{x}_t(\epsilon)) - \mathbf{v}[f,\epsilon]_t^N(\epsilon) \right) + \sum_{t=1}^{T} \epsilon_t (\sum_{j=1}^{N} \left(\mathbf{v}[f,\epsilon]_t^j(\epsilon) - \mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon) \right)) \right] \\
&\leq \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \epsilon_t \left(f\left(\mathbf{x}_t(\epsilon)\right) - \mathbf{v}[f,\epsilon]_t^N(\epsilon) \right) \right] \\
&+ \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \epsilon_t \left(\sum_{j=1}^{N} \left(\mathbf{v}[f,\epsilon]_t^j(\epsilon) - \mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon) \right) \right) \right]
\end{aligned}
\tag{4}
$$

The first term above can be bounded via Cauchy-Schwarz inequality as

$$
\begin{aligned}
&\mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \epsilon_t \left(f\left(\mathbf{x}_t(\epsilon)\right) - \mathbf{v}[f,\epsilon]_t^N(\epsilon) \right) \right] \\
&\leq T\mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \frac{\epsilon_t}{\sqrt{T}} \frac{\left(f\left(\mathbf{x}_t(\epsilon)\right) - \mathbf{v}[f,\epsilon]_t^N(\epsilon) \right)}{\sqrt{T}} \right] \leq T\beta_N.
\end{aligned}
$$

Proof of the Theorem (3)

The second term in (4) is bounded by considering successive refinements of the cover. Consider all possible pairs of $\mathbf{v}^s \in V_j$ and $\mathbf{v}^r \in V_{j-1}$, for $1 \leq s \leq |V_j|, 1 \leq r \leq |V_{j-1}|$, where we assumed an arbitrary enumeration of elements. For each pair $(\mathbf{v}^s, \mathbf{v}^r)$, define a real-valued tree $\mathbf{w}^{(s,r)}$ by

$$\mathbf{w}_t^{(s,r)}(\epsilon) = \begin{cases} \mathbf{v}_t^s(\epsilon) - \mathbf{v}_t^r(\epsilon) & \text{if there exists } f \in \mathcal{F} \text{ s.t. } \mathbf{v}^s = \mathbf{v}[f,\epsilon]^j, \mathbf{v}^r = \mathbf{v}[f,\epsilon]^{j-1} \\ 0 & \text{otherwise.} \end{cases}$$

for all $t \in [T]$ and $\epsilon \in \pm 1^T$. It is crucial that $\mathbf{w}_t^{(s,r)}$ can be *non-zero only on those paths ϵ for which $\mathbf{v}^s$ and $\mathbf{v}^r$ are indeed the members of the covers (at successive resolutions) close to $f(\mathbf{x}(\epsilon))$ (in the ℓ_2 sense) for some path $f \in \mathcal{F}$.* Let the set of trees W_j be defined as

$$W_j = \left\{ \mathbf{w}^{(s,r)} : 1 \leq s \leq |V_j|, 1 \leq r \leq |V_{j-1}| \right\}$$

Proof of the Theorem (4)

Now, the second term in (4) can be written as

$$\begin{aligned}
&\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^{T}\epsilon_t\sum_{j=1}^{N}\left(\mathbf{v}[f,\epsilon]_t^j(\epsilon)-\mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon)\right)\right]\\
&\le\sum_{j=1}^{N}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^{T}\epsilon_t\left(\mathbf{v}[f,\epsilon]_t^j(\epsilon)-\mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon)\right)\right]\\
&\le\sum_{j=1}^{N}\mathbb{E}_\epsilon\left[\max_{\mathbf{w}\in W_j}\sum_{t=1}^{T}\epsilon_t\mathbf{w}_t(\epsilon)\right]
\end{aligned}$$

The last inequality holds because for any $j\in[N],\epsilon\in\pm1^F$ and $f\in\mathcal{F}$ there is sime $\mathbf{w}^{(s,r)}\in W_j$ with $\mathbf{v}[f,\epsilon]^j=\mathbf{s},\mathbf{v}[f,\epsilon]^{j-1}=\mathbf{r}$ and

$$\mathbf{v}_t^s(\epsilon)-\mathbf{v}_t^r(\epsilon)=\mathbf{w}_t^{(s,r)}(\epsilon),\forall t\le T$$

clearly, $|W_j|\le|V_j||V_{j-1}|$.

Proof of the Theorem (5)

To invoke Lemma, it remains to bound the magnitude of all $\mathbf{w}^{(s,r)} \in W_j$ along all paths. For this purpose, fix $\mathbf{w}^{(s,r)}$ and a path ϵ.

Condition 1

If there exists $f \in \mathcal{F}$ for which $\mathbf{v}^s = \mathbf{v}[f,\epsilon]^j$ and $\mathbf{v}^r = \mathbf{v}[f,\epsilon]^{j-1}$, then $\mathbf{w}_t^{(s,r)}(\epsilon) = \mathbf{v}[f,\epsilon]_t^j - \mathbf{v}[f,\epsilon]_t^{j-1}$ for any $t \in [T]$. By triangle inequality

$$\begin{aligned}\sqrt{\sum_{t=1}^{T} \mathbf{w}_t^{(s,r)}(\epsilon)^2} &\leq \sqrt{\sum_{t=1}^{T} \left(\mathbf{v}[f,\epsilon]_t^j(\epsilon) - f(\mathbf{x}_t(\epsilon))\right)^2} + \sqrt{\sum_{t=1}^{T} \left(\mathbf{v}[f,\epsilon]_t^{j-1}(\epsilon) - f\left(\mathbf{x}_t(\epsilon)\right)\right)^2} \\ &\leq \sqrt{T}\left(\beta_j + \beta_{j-1}\right) = 3\sqrt{T}\beta_j\end{aligned}$$

Proof of the Theorem (6)

Condition 2

If there exists no such $f \in \mathcal{F}$ for the given ϵ and (s, r), then $\mathbf{w}_t^{(s,r)}(\epsilon)$ is zero for all $t \geq t_o$, for sime $1 \leq t_o \leq T$, and thus

$$\sqrt{\sum_{t=1}^{T} \mathbf{w}_t^{(s,r)}(\epsilon)^2} \leq \sqrt{\sum_{t=1}^{T} \mathbf{w}_t^{(s,r)}\left(\epsilon'\right)^2}$$

for any other path ϵ' which agrees with ϵ up to t_o. Hence, the bound

$$\sqrt{\sum_{t=1}^{T} \mathbf{w}_t^{(s,r)}(\epsilon)^2} \leq 3\sqrt{T}\beta_j$$

holds for all $\epsilon \in \{\pm 1\}^T$ and $\mathbf{w}^{(s,r)} \in W_j$

Proof of the Theorem (7)

Lemma 1

For any finite set V of $\mathbb{R}$-valued trees of depth T we have that

$$\mathbb{E}_\epsilon \left[\max_{\mathbf{v} \in V} \sum_{t=1}^{T} \epsilon_t \mathbf{v}_t(\epsilon) \right] \leq \sqrt{2 \log(|V|) \max_{\mathbf{v} \in V} \max_{\mathbf{v} \in V} \max_{\epsilon \in \{\pm 1\}^T} \sum_{t=1}^{T} \mathbf{v}_t(\epsilon)^2}$$

Back to Equation (4)

We put everything together and apply Lemma:

$$\begin{aligned}
\mathbb{E}_\epsilon \left[\sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \epsilon_t f\left(\mathbf{x}_t(\epsilon)\right) \right] &\leq T\beta_N + \sqrt{T} \sum_{j=1}^{N} 3\beta_j \sqrt{2 \log\left(|V_j|\,|V_{j-1}|\right)} \\
&\leq T\beta_N + \sqrt{T} \sum_{j=1}^{N} 6\beta_j \sqrt{\log(|V_j|)} \\
&\leq T\beta_N + 12\sqrt{T} \sum^{N} \left(\beta_j - \beta_{j+1}\right) \sqrt{\log \mathcal{N}_2\left(\beta_j, \mathcal{F}, \mathbf{x}\right)}
\end{aligned}$$

Proof of the Theorem (8)

$$\leq T\beta_N + 12\int_{\beta_{N+1}}^{\beta_0} \sqrt{T \log \mathcal{N}_2(\delta, \mathcal{F}, \mathbf{x})}d\delta$$

where the last but one step is because $2(\beta_j - \beta_{j+1}) = \beta_j$. Now for any $\alpha > 0$, pick $N = \sup\{j : \beta_j > 2\alpha\}$. In this case we see that by our choice of N, $\beta_{N+1} < 2\alpha$ and so $\beta_N = 2\beta_{N+1} \leq 4\alpha$. Also note that since $\beta_N > 2\alpha, \beta_{N+1} = \frac{\beta_N}{2} > \alpha$. Hence we conclude that

$$\mathfrak{R}_T(\mathcal{F}) \leq \inf_{\alpha}\left\{4T\alpha + 12\int_{\alpha}^{1} \sqrt{T \log \mathcal{N}_2(\delta, \mathcal{F}, T)}d\delta\right\} = \mathfrak{D}_T(\mathcal{F})$$