

We continue to prove the lemma from Lecture 18:

Lemma 19.1. *Let $\mathcal{F}_d = \text{conv}_d \mathcal{H} = \{\sum_{i=1}^d \lambda_i h_i, h_i \in \mathcal{H}\}$ and fix $\delta \in (0, 1]$. Then*

$$\mathbb{P} \left(\forall f \in \mathcal{F}_d, \frac{\mathbb{E}\varphi_\delta - \bar{\varphi}_\delta}{\sqrt{\mathbb{E}\varphi_\delta}} \leq K \left(\sqrt{\frac{dV \log \frac{n}{\delta}}{n}} + \sqrt{\frac{t}{n}} \right) \right) \geq 1 - e^{-t}.$$

Proof. We showed that

$$\log \mathcal{D}(\varphi_\delta(y\mathcal{F}_d), \varepsilon/2, d_{x,y}) \leq KdV \log \frac{2}{\varepsilon\delta}.$$

By the result of Lecture 16,

$$\mathbb{E}\varphi_\delta(yf(x)) - \frac{1}{n} \sum_{i=1}^n \varphi_\delta(y_i f(x_i)) \leq \frac{k}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}\varphi_\delta}} \log^{1/2} \mathcal{D}(\varphi_\delta(y\mathcal{F}_d(x)), \varepsilon) d\varepsilon + \sqrt{\frac{t\mathbb{E}\varphi_\delta}{n}}$$

with probability at least $1 - e^{-t}$. We have

$$\begin{aligned} \frac{k}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}\varphi_\delta}} \log^{1/2} \mathcal{D}(\varphi_\delta(y\mathcal{F}_d(x)), \varepsilon) d\varepsilon &\leq \frac{k}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}\varphi_\delta}} \sqrt{dV \log \frac{2}{\varepsilon\delta}} d\varepsilon \\ &= \frac{k}{\sqrt{n}} \frac{2}{\delta} \int_0^{\delta\sqrt{\mathbb{E}\varphi_\delta/2}} \sqrt{dV} \sqrt{\log \frac{1}{x}} dx \\ &\leq \frac{k}{\sqrt{n}} \frac{2}{\delta} \sqrt{dV} 2 \frac{\delta}{2} \sqrt{\mathbb{E}\varphi_\delta} \sqrt{\log \frac{2}{\delta\sqrt{\mathbb{E}\varphi_\delta}}} \end{aligned}$$

where we have made a change of variables $\frac{2}{\varepsilon\delta} = x$, $\varepsilon = \frac{2x}{\delta}$. Without loss of generality, assume $\mathbb{E}\varphi_\delta \geq 1/n$. Otherwise, we're doing better than in Lemma: $\frac{\mathbb{E}}{\sqrt{\mathbb{E}}} \leq \sqrt{\frac{\log n}{n}} \Rightarrow \mathbb{E} \leq \frac{\log n}{n}$. Hence,

$$\begin{aligned} \frac{k}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}\varphi_\delta}} \log^{1/2} \mathcal{D}(\varphi_\delta(y\mathcal{F}_d(x)), \varepsilon) d\varepsilon &\leq K \sqrt{\frac{dV\mathbb{E}\varphi_\delta}{n} \log \frac{2\sqrt{n}}{\delta}} \\ &\leq K \sqrt{\frac{dV\mathbb{E}\varphi_\delta}{n} \log \frac{n}{\delta}} \end{aligned}$$

So, with probability at least $1 - e^{-t}$,

$$\mathbb{E}\varphi_\delta(yf(x)) - \frac{1}{n} \sum_{i=1}^n \varphi_\delta(y_i f(x_i)) \leq K \sqrt{\frac{dV\mathbb{E}\varphi_\delta(yf(x))}{n} \log \frac{n}{\delta}} + \sqrt{\frac{t\mathbb{E}\varphi_\delta(yf(x))}{n}}$$

which concludes the proof. □

The above lemma gives a result for a fixed $d \geq 1$ and $\delta \in (0, 1]$. To obtain a uniform result, it's enough to consider $\delta \in \Delta = \{2^{-k}, k \geq 1\}$ and $d \in \{1, 2, \dots\}$. For a fixed δ and d , use the Lemma above with $t_{\delta,d}$ defined by $e^{-t_{\delta,d}} = e^{-t} \frac{6\delta}{d^2\pi^2}$. Then

$$\mathbb{P} \left(\forall f \in \mathcal{F}_d, \dots + \sqrt{\frac{t_{\delta,d}}{n}} \right) \geq 1 - e^{-t_{\delta,d}} = 1 - e^{-t} \frac{6\delta}{d^2\pi^2}$$

and

$$\mathbb{P} \left(\bigcup_{d,\delta} \left\{ \forall f \in \mathcal{F}_d, \dots + \sqrt{\frac{t_{\delta,d}}{n}} \right\} \right) \geq 1 - \sum_{\delta,d} e^{-t} \frac{6\delta}{d^2\pi^2} = 1 - e^{-t}.$$

Since $t_{\delta,d} = t + \log \frac{d^2\pi^2}{6\delta}$,

$$\begin{aligned} \forall f \in \mathcal{F}_d, \frac{\mathbb{E}\varphi_\delta - \bar{\varphi}_\delta}{\sqrt{\mathbb{E}\varphi_\delta}} &\leq K \left(\sqrt{\frac{dV \log \frac{n}{\delta}}{n}} + \sqrt{\frac{t + \log \frac{d^2\pi^2}{6\delta}}{n}} \right) \\ &\leq K \left(\sqrt{\frac{dV \log \frac{n}{\delta}}{n}} + \sqrt{\log \frac{d^2\pi^2}{6\delta}} + \sqrt{\frac{t}{n}} \right) \\ &\leq K' \left(\sqrt{\frac{dV \log \frac{n}{\delta}}{n}} + \sqrt{\frac{t}{n}} \right) \end{aligned}$$

since $\log \frac{d^2\pi^2}{6\delta}$, the penalty for union-bound, is much smaller than $\sqrt{\frac{dV \log \frac{n}{\delta}}{n}}$.

Recall the bound on the misclassification error

$$\mathbb{P}(yf(x) \leq 0) \leq \frac{1}{n} \sum_{i=1}^n I(y_i f(x_i) \leq \delta) + \left(\mathbb{E}\varphi_\delta(yf(x)) - \frac{1}{n} \sum_{i=1}^n \varphi_\delta(y_i f(x_i)) \right).$$

If

$$\frac{\mathbb{E}\varphi_\delta - \frac{1}{n} \sum_{i=1}^n \varphi_\delta}{\sqrt{\mathbb{E}\varphi_\delta}} \leq \varepsilon,$$

then

$$\mathbb{E}\varphi_\delta - \varepsilon \sqrt{\mathbb{E}\varphi_\delta} - \frac{1}{n} \sum_{i=1}^n \varphi_\delta \leq 0.$$

Hence,

$$\sqrt{\mathbb{E}\varphi_\delta} \leq \frac{\varepsilon}{2} + \sqrt{\left(\frac{\varepsilon}{2}\right)^2 + \frac{1}{n} \sum_{i=1}^n \varphi_\delta}$$

$$\mathbb{E}\varphi_\delta \leq 2\left(\frac{\varepsilon}{2}\right)^2 + 2\frac{1}{n} \sum_{i=1}^n \varphi_\delta.$$

The bound becomes

$$\mathbb{P}(yf(x) \leq 0) \leq K \left(\frac{1}{n} \sum_{i=1}^n I(y_i f(x_i) \leq \delta) + \underbrace{\frac{dV}{n} \log \frac{n}{\delta}}_{(*)} + \frac{t}{n} \right)$$

where K is a rough constant.

(*) not satisfactory because in boosting the bound should get better when the number of functions grows. We prove a better bound in the next lecture.