In Lecture 24 we proved

$$\mathbb{E} \sup_{h \in \mathcal{H}_k(A_1,\ldots,A_k)} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (y_i - h(x_i))^2 \right| \le 8 \prod_{j=1}^k (2LA_j) \cdot \mathbb{E} \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i h(x_i) \right| + \frac{8}{\sqrt{n}}$$

Hence,

$$\begin{aligned} Z\left(\mathcal{H}_k(A_1,\ldots,A_k)\right) &:= \sup_{h \in \mathcal{H}_k(A_1,\ldots,A_k)} \left| \mathbb{E}\mathcal{L}(y, h(x)) - \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y_i, h(x_i)) \right| \\ &\le 8 \prod_{j=1}^k (2LA_j) \cdot \mathbb{E} \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i h(x_i) \right| + \frac{8}{\sqrt{n}} + 8\sqrt{\frac{t}{n}} \end{aligned}$$

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

Assume $\mathcal{H}$ is a VC-subgraph class, $-1 \le h \le 1$.

We had the following result:

$$\begin{aligned} \mathbb{P}_\varepsilon \Bigg(\forall h \in \mathcal{H},\ \frac{1}{n} \sum_{i=1}^n \varepsilon_i h(x_i) \le \frac{K}{\sqrt{n}} \int_0^{\sqrt{\frac{1}{n}\sum_{i=1}^n h^2(x_i)}} \log^{1/2} \mathcal{D}(\mathcal{H}, \varepsilon, d_x) d\varepsilon \\ + K \sqrt{\frac{t}{n} \left(\frac{1}{n} \sum_{i=1}^n h^2(x_i) \right)} \Bigg) \ge 1 - e^{-t}, \end{aligned}$$

where

$$d_x(f, g) = \left(\frac{1}{n} \sum_{i=1}^n (f(x_i) - g(x_i))^2 \right)^{1/2}.$$

Furthermore,

$$\begin{aligned} \mathbb{P}_\varepsilon \Bigg(\forall h \in \mathcal{H},\ \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i h(x_i) \right| \le \frac{K}{\sqrt{n}} \int_0^{\sqrt{\frac{1}{n}\sum_{i=1}^n h^2(x_i)}} \log^{1/2} \mathcal{D}(\mathcal{H}, \varepsilon, d_x) d\varepsilon \\ + K \sqrt{\frac{t}{n} \left(\frac{1}{n} \sum_{i=1}^n h^2(x_i) \right)} \Bigg) \ge 1 - 2e^{-t}, \end{aligned}$$

Since $-1 \le h \le 1$ for all $h \in \mathcal{H}$,

$$\mathbb{P}_\varepsilon \left(\sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i h(x_i) \right| \le \frac{K}{\sqrt{n}} \int_0^1 \log^{1/2} \mathcal{D}(\mathcal{H}, \varepsilon, d_x) d\varepsilon + K \sqrt{\frac{t}{n}} \right) \ge 1 - 2e^{-t},$$

Since $\mathcal{H}$ is a VC-subgraph class with $VC(\mathcal{H}) = V$,

$$\log \mathcal{D}(\mathcal{H}, \varepsilon, d_x) \leq KV \log \frac{2}{\varepsilon}.$$

Hence,

$$\begin{aligned} \int_0^1 \log^{1/2} \mathcal{D}(\mathcal{H}, \varepsilon, d_x) d\varepsilon &\leq \int_0^1 \sqrt{KV \log \frac{2}{\varepsilon}} d\varepsilon \\ &\leq K\sqrt{V} \int_0^1 \sqrt{\log \frac{2}{\varepsilon}} d\varepsilon \leq K\sqrt{V} \end{aligned}$$

Let $\xi \geq 0$ be a random variable. Then

$$\begin{aligned} \mathbb{E}\xi &= \int_0^\infty \mathbb{P}(\xi \geq t)\, dt = \int_0^a \mathbb{P}(\xi \geq t)\, dt + \int_a^\infty \mathbb{P}(\xi \geq t)\, dt \\ &\leq a + \int_a^\infty \mathbb{P}(\xi \geq t)\, dt = a + \int_0^\infty \mathbb{P}(\xi \geq a + u)\, du \end{aligned}$$

Let $K\sqrt{\frac{V}{n}} = a$ and $K\sqrt{\frac{t}{n}} = u$. Then $e^{-t} = e^{-\frac{nu^2}{K^2}}$. Hence, we have

$$\begin{aligned} \mathbb{E}_\varepsilon \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i h(x_i) \right| &\leq K\sqrt{\frac{V}{n}} + \int_0^\infty 2e^{-\frac{nu^2}{K^2}} du \\ &= K\sqrt{\frac{V}{n}} + \int_0^\infty \frac{K}{\sqrt{n}} e^{-x^2} dx \\ &\leq K\sqrt{\frac{V}{n}} + \frac{K}{\sqrt{n}} \leq K\sqrt{\frac{V}{n}} \end{aligned}$$

for $V \geq 2$. We made a change of variable so that $x^2 = \frac{nu^2}{K^2}$. Constants K change their values from line to line.

We obtain,

$$Z\left(\mathcal{H}_k(A_1, \ldots, A_k)\right) \leq K \prod_{j=1}^k (2LA_j) \cdot \sqrt{\frac{V}{n}} + \frac{8}{\sqrt{n}} + 8\sqrt{\frac{t}{n}}$$

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

Assume that for any j, $A_j \in (2^{-\ell_j - 1}, 2^{-\ell_j}]$. This defines ℓ_j. Let

$$\mathcal{H}_k(\ell_1, \ldots, \ell_k) = \bigcup \left\{ \mathcal{H}_k(A_1, \ldots, A_k) : A_j \in (2^{-\ell_j - 1}, 2^{-\ell_j}] \right\}.$$

Then the empirical process

$$Z\left(\mathcal{H}_k(\ell_1,\ldots,\ell_k)\right) \le K\prod_{j=1}^{k}(2L\cdot 2^{-\ell_j})\cdot\sqrt{\frac{V}{n}}+\frac{8}{\sqrt{n}}+8\sqrt{\frac{t}{n}}$$

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

For a given sequence $(\ell_1,\ldots,\ell_k)$, redefine t as $t+2\sum_{j=1}^k \log|w_j|$ where $w_j=\ell_j$ if $\ell_j\neq 0$ and $w_j=1$ if $\ell_j=0$.

With this t,

$$Z\left(\mathcal{H}_k(\ell_1,\ldots,\ell_k)\right) \le K\prod_{j=1}^{k}(2L\cdot 2^{-\ell_j})\cdot\sqrt{\frac{V}{n}}+\frac{8}{\sqrt{n}}+8\sqrt{\frac{t+2\sum_{j=1}^k\log|w_j|}{n}}$$

with probability at least

$$1-e^{-t-2\sum_{j=1}^k\log|w_j|}=1-\prod_{j=1}^k\frac{1}{|w_j|^2}e^{-t}.$$

By union bound, the above holds for all $\ell_1,\ldots,\ell_k\in\mathcal{Z}$ with probability at least

$$\begin{aligned}1-\sum_{\ell_1,\ldots,\ell_k\in\mathcal{Z}}\prod_{j=1}^k\frac{1}{|w_j|^2}e^{-t} &= 1-\left(\sum_{\ell_1\in\mathcal{Z}}\frac{1}{|w_1|^2}\right)^k e^{-t}\\ &= 1-\left(1+2\frac{\pi^2}{6}\right)^k e^{-t}\ge 1-5^k e^{-t}=1-e^{-u}\end{aligned}$$

for $t=u+k\log 5$.

Hence, with probability at least $1-e^{-u}$,

$$\forall(\ell_1,\ldots,\ell_k),\ Z\left(\mathcal{H}_k(\ell_1,\ldots,\ell_k)\right)\le K\prod_{j=1}^k(2L\cdot 2^{-\ell_j})\cdot\sqrt{\frac{V}{n}}+\frac{8}{\sqrt{n}}+8\sqrt{\frac{2\sum_{j=1}^k\log|w_j|+k\log 5+u}{n}}.$$

If $A_j\in(2^{-\ell_j-1},2^{-\ell_j}]$, then $-\ell_j-1\le\log A_j\le\ell_j$ and $|\ell_j|\le|\log A_j|+1$. Hence, $|w_j|\le|\log A_j|+1$. Therefore, with probability at least $1-e^{-u}$,

$$\begin{aligned}\forall(\mathcal{A}_1,\ldots,A_k),\ Z\left(\mathcal{H}_k(A_1,\ldots,A_k)\right)\le K\prod_{j=1}^k(4L\cdot A_j)\cdot\sqrt{\frac{V}{n}}+\frac{8}{\sqrt{n}}\\ +8\sqrt{\frac{2\sum_{j=1}^k\log\left(|\log A_j|+1\right)+k\log 5+u}{n}}.\end{aligned}$$

Notice that $\log(|\log A_j| + 1)$ is large when A_j is very large or very small. This is penalty and we want the product term to be dominating. But $\log \log A_j \leq 5$ for most practical applications.