Let $f = \sum_{i=1}^{T} \lambda_i h_i$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_T \geq 0$. Rewrite f as

$$f = \sum_{i=1}^{d} \lambda_i h_i + \sum_{i=d+1}^{T} \lambda_i h_i = \sum_{i=1}^{d} \lambda_i h_i + \gamma(d) \sum_{i=d+1}^{T} \lambda_i' h_i$$

where $\gamma(d) = \sum_{i=d+1}^{T} \lambda_i$ and $\lambda_i' = \lambda_i / \gamma(d)$.

Consider the following random approximation of f,

$$g = \sum_{i=1}^{d} \lambda_i h_i + \gamma(d) \frac{1}{k} \sum_{j=1}^{k} Y_j$$

where, as in the previous lectures,

$$\mathbb{P}\left(Y_j = h_i\right) = \lambda_i', \quad i = d+1, \ldots, T$$

for any $j = 1, \ldots, k$. Recall that $\mathbb{E} Y_j = \sum_{i=d+1}^{T} \lambda_i' h_i$.

Then

$$\begin{aligned} \mathbb{P}\left(yf(x) \leq 0\right) &= \mathbb{P}\left(yf(x) \leq 0, yg(x) \leq \delta\right) + \mathbb{P}\left(yf(x) \leq 0, yg(x) > \delta\right) \\ &\leq \mathbb{P}\left(yg(x) \leq \delta\right) + \mathbb{E}\left[\mathbb{P}_Y\left(yf(x) \leq 0, yg(x) \geq \delta \;\middle|\; (x,y)\right)\right] \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathbb{P}_Y\left(yf(x) \leq 0, yg(x) \geq \delta \;\middle|\; (x,y)\right) &\leq \mathbb{P}_Y\left(yg(x) - yf(x) > \delta \;\middle|\; (x,y)\right) \\ &= \mathbb{P}_Y\left(\gamma(d) y \left(\frac{1}{k}\sum_{j=1}^{k} Y_j(x) - \mathbb{E} Y_1\right) \geq \delta \;\middle|\; (x,y)\right). \end{aligned}$$

By renaming $Y_j' = \frac{yY_j + 1}{2} \in [0,1]$ and applying Hoeffding's inequality, we get

$$\begin{aligned} \mathbb{P}_Y\left(\gamma(d) y \left(\frac{1}{k}\sum_{j=1}^{k} Y_j(x) - \mathbb{E} Y\right) \geq \delta \;\middle|\; (x,y)\right) &= \mathbb{P}_Y\left(\frac{1}{k}\sum_{j=1}^{k} Y_j'(x) - \mathbb{E} Y_1' \geq \frac{\delta}{2\gamma(d)} \;\middle|\; (x,y)\right) \\ &\leq e^{-\frac{k\delta^2}{2\gamma(d)^2}}. \end{aligned}$$

Hence,

$$\mathbb{P}\left(yf(x) \leq 0\right) \leq \mathbb{P}\left(yg(x) \leq \delta\right) + e^{-\frac{k\delta^2}{2\gamma^2(d)}}.$$

If we set $e^{-\frac{k\delta^2}{2\gamma(d)^2}} = \frac{1}{n}$, then $k = \frac{2\gamma^2(d)}{\delta^2}\log n$.

We have

$$g = \sum_{i=1}^{d} \lambda_i h_i + \gamma(d)\frac{1}{k}\sum_{j=1}^{k} Y_j \in \mathrm{conv}_{d+k}\mathcal{H},$$

$d + k = d + \frac{2\gamma^2(d)}{\delta^2}\log n$.

Define the effective dimension of f as

$$e(f,\delta) = \min_{0\le d\le T}\left(d + \frac{2\gamma^2(d)}{\delta^2}\log n\right).$$

Recall from the previous lectures that

$$\mathbb{P}_n\left(yg(x) \le 2\delta\right) \le \mathbb{P}_n\left(yf(x) \le 3\delta\right) + \frac{1}{n}.$$

Hence, we have the following *margin-sparsity bound*

Theorem 21.1. *For $\lambda_1 \ge \ldots \lambda_T \ge 0$, we define $\gamma(d,f) = \sum_{i=d+1}^{T}\lambda_i$. Then with probability at least $1-e^{-t}$,*

$$\mathbb{P}\left(yf(x) \le 0\right) \le \inf_{\delta\in(0,1)}\left(\varepsilon + \sqrt{\mathbb{P}_n\left(yf(x)\le\delta\right) + \varepsilon^2}\right)^2$$

where

$$\varepsilon = K\left(\sqrt{\frac{V\cdot e(f,\delta)}{n}\log\frac{n}{\delta}} + \sqrt{\frac{t}{n}}\right)$$

Example 1. *Consider the zero-error case. Define*

$$\delta^* = \sup\{\delta > 0, \mathbb{P}_n\left(yf(x)\le\delta\right) = 0\}.$$

Hence, $\mathbb{P}_n\left(yf(x)\le\delta^*\right) = 0$ for confidence δ^*. Then

$$\begin{aligned}\mathbb{P}\left(yf(x)\le 0\right) \le 4\varepsilon^2 &= K\left(\frac{V\cdot e(f,\delta^*)}{n}\log\frac{n}{\delta^*} + \frac{t}{n}\right)\\ &\le K\left(\frac{V\log n}{(\delta^*)^2 n}\log\frac{n}{\delta^*} + \frac{t}{n}\right)\end{aligned}$$

because $e(f,\delta) \le \frac{2}{\delta^2}\log n$ always.

Example 2. *Consider the polynomial weight decay: $\lambda_i \leq K i^{-\alpha}$, for some $\alpha > 1$. Then*

$$\gamma(d) = \sum_{i=d+1}^{T} \lambda_i \leq K \sum_{i=d+1}^{T} i^{-\alpha} \leq K \int_d^\infty x^{-\alpha} dx = K \frac{1}{(\alpha-1)d^{\alpha-1}} = \frac{K_\alpha}{d^{\alpha-1}}$$

Then

$$\begin{aligned} e(f,\delta) &= \min_d \left(d + \frac{2\gamma^2(d)}{\delta^2} \log n \right) \\ &\leq \min_d \left(d + \frac{K'_\alpha}{\delta^2 d^{2(\alpha-1)}} \log n \right) \end{aligned}$$

Taking derivative with respect to d and setting it to zero,

$$1 - \frac{K_\alpha \log n}{\delta^2 d^{2\alpha-1}} = 0$$

we get

$$d = K_\alpha \cdot \frac{\log^{1/(2\alpha-1)} n}{\delta^{2/(2\alpha-1)}} \leq K \frac{\log n}{\delta^{2/(2\alpha-1)}}.$$

Hence,

$$e(f,\delta) \leq K \frac{\log n}{\delta^{2/(2\alpha-1)}}$$

Plugging in,

$$\mathbb{P}\left(yf(x) \leq 0\right) \leq K \left(\frac{V \log n}{n(\delta^*)^{2/(2\alpha-1)}} \log \frac{n}{\delta^*} + \frac{t}{n} \right).$$

As $\alpha \to \infty$, the bound behaves like

$$\frac{V \log n}{n} \log \frac{n}{\delta^*}.$$