Theorem 22.1. *With probability at least $1-e^{-t}$, for any $T \geq 1$ and any $f = \sum_{i=1}^{T} \lambda_i h_i$,*

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

where $\varepsilon = \varepsilon(\delta) = K\left(\sqrt{\frac{V \min(T,(\log n)/\delta^2) \log \frac{n}{\delta}}{n}} + \sqrt{\frac{t}{n}}\right)$.

Here we used the notation $\mathbb{P}_n\left(C\right) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \in C)$.

Remark:

$$\mathbb{P}\left(yf(x) \leq 0\right) \leq \inf_{\delta \in (0,1)} K\left(\underbrace{\mathbb{P}_n\left(yf(x) \leq \delta\right)}_{\text{inc. with } \delta} + \underbrace{\frac{V \min(T,(\log n)/\delta^2) \log \frac{n}{\delta}}{n}}_{\text{dec. with } \delta} + \frac{t}{n}\right).$$

Proof. Let $f = \sum_{i=1}^{T} \lambda_i h_i$, $g = \frac{1}{k}\sum_{j=1}^{k} Y_j$, where

$$\mathbb{P}\left(Y_j = h_i\right) = \lambda_i \ \text{ and } \ \mathbb{P}\left(Y_j = 0\right) = 1 - \sum_{i=1}^{T} \lambda_i$$

as in Lecture 17. Then $\mathbb{E}Y_j(x) = f(x)$.

$$\begin{aligned}\mathbb{P}\left(yf(x) \leq 0\right) &= \mathbb{P}\left(yf(x) \leq 0, yg(x) \leq \delta\right) + \mathbb{P}\left(yf(x) \leq 0, yg(x) > \delta\right)\\ &\leq \mathbb{P}\left(yg(x) \leq \delta\right) + \mathbb{P}\left(yg(x) > \delta \mid yf(x) \leq 0\right)\end{aligned}$$

$$\mathbb{P}\left(yg(x) > \delta \;\middle|\; yf(x) \leq 0\right) = \mathbb{E}_x \mathbb{P}_Y\left(y\frac{1}{k}\sum_{j=1}^{k} Y_j(x) > \delta \;\middle|\; y\mathbb{E}_Y Y_j(x) \leq 0\right)$$

Shift Y's to $[0,1]$ by defining $Y'_j = \frac{yY_j + 1}{2}$. Then

$$\begin{aligned}\mathbb{P}\left(yg(x) > \delta | yf(x) \leq 0\right) &= \mathbb{E}_x \mathbb{P}_Y\left(\frac{1}{k}\sum_{j=1}^{k} Y'_j \geq \frac{1}{2} + \frac{\delta}{2} \;\middle|\; \mathbb{E}Y'_j \leq \frac{1}{2}\right)\\ &\leq \mathbb{E}_x \mathbb{P}_Y\left(\frac{1}{k}\sum_{j=1}^{k} Y'_j \geq \mathbb{E}Y'_1 + \frac{\delta}{2} \;\middle|\; \mathbb{E}Y'_j \leq \frac{1}{2}\right)\\ &\leq \text{(by Hoeffding's ineq.) } \mathbb{E}_x e^{-kD\left(\mathbb{E}Y'_1 + \frac{\delta}{2}, \mathbb{E}Y'_1\right)}\\ &\leq \mathbb{E}_x e^{-k\delta^2/2} = e^{-k\delta^2/2}\end{aligned}$$

because $D(p,q) \geq 2(p-q)^2$ (KL-divergence for binomial variables, Homework 1) and, hence,

$$D\left(\mathbb{E}Y'_1 + \frac{\delta}{2}, \mathbb{E}Y'_1\right) \geq 2\left(\frac{\delta}{2}\right)^2 = \delta^2/2.$$

We therefore obtain

$$\mathbb{P}\left(yf(x) \leq 0\right) \leq \mathbb{P}\left(yg(x) \leq \delta\right) + e^{-k\delta^2/2} \tag{22.1}$$

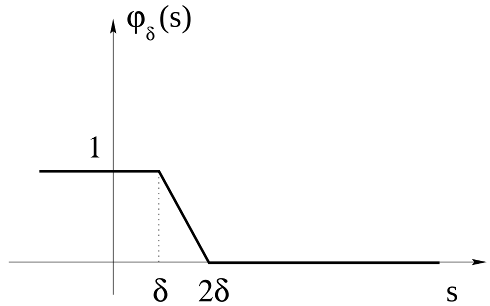
and the second term in the bound will be chosen to be equal to $1/n$.

Similarly, we can show

$$\mathbb{P}_n\left(yg(x) \leq 2\delta\right) \leq \mathbb{P}_n\left(yf(x) \leq 3\delta\right) + e^{-k\delta^2/2}.$$

Choose k such that $e^{-k\delta^2/2} = 1/n$, i.e. $k = \frac{2}{\delta^2}\log n$.

Now define φ_δ as follows:



Observe that

$$(22.2) \qquad I(s \leq \delta) \leq \varphi_\delta\left(s\right) \leq I(s \leq 2\delta).$$

By the result of Lecture 21, with probability at least $1 - e^{-t}$, for all k, δ and any $g \in \mathcal{F}_k = \text{conv}_{\ k}(\mathcal{H})$,

$$\begin{aligned}\Phi\left(\mathbb{E}\varphi_\delta, \frac{1}{n}\sum_{i=1}^n \varphi_\delta\right) &= \frac{\mathbb{E}\varphi_\delta\left(yg(x)\right) - \frac{1}{n}\sum_{i=1}^n \varphi_\delta\left(y_i g(x_i)\right)}{\sqrt{\mathbb{E}\varphi_\delta\left(yg(x)\right)}} \\ &\leq K\left(\sqrt{\frac{Vk\log\frac{n}{\delta}}{n}} + \sqrt{\frac{t}{n}}\right) \\ &= \varepsilon/2.\end{aligned}$$

Note that $\Phi(x, y) = \frac{x-y}{\sqrt{x}}$ is increasing with x and decreasing with y.

By inequalities (22.1) and (22.2),

$$\mathbb{E}\varphi_\delta\left(yg(x)\right) \geq \mathbb{P}\left(yg(x) \leq \delta\right) \geq \mathbb{P}\left(yf(x) \leq 0\right) - \frac{1}{n}$$

and

$$\frac{1}{n}\sum_{i=1}^n \varphi_\delta\left(y_i g(x_i)\right) \leq \mathbb{P}_n\left(yg(x) \leq 2\delta\right) \leq \mathbb{P}_n\left(yf(x) \leq 3\delta\right) + \frac{1}{n}.$$

By decreasing x and increasing y in $\Phi(x, y)$, we decrease $\Phi(x, y)$. Hence,

$$\Phi\left(\underbrace{\mathbb{P}\left(yf(x) \leq 0\right) - \frac{1}{n}}_{x}, \underbrace{\mathbb{P}_n\left(yf(x) \leq 3\delta\right) + \frac{1}{n}}_{y}\right) \leq K\left(\sqrt{\frac{Vk\log\frac{n}{\delta}}{n}} + \sqrt{\frac{t}{n}}\right)$$

where $k = \frac{2}{\delta^2}\log n$.

If $\frac{x-y}{\sqrt{x}} \leq \varepsilon$, we have

$$x \leq \left(\frac{\varepsilon}{2} + \sqrt{\left(\frac{\varepsilon}{2}\right)^2 + y}\right)^2$$

So,

$$\mathbb{P}\left(yf(x) \leq 0\right) - \frac{1}{n} \leq \left(\frac{\varepsilon}{2} + \sqrt{\left(\frac{\varepsilon}{2}\right)^2 + \mathbb{P}_n\left(yf(x) \leq 3\delta\right) + \frac{1}{n}}\right)^2.$$

□