Consider the classification setting, i.e. $\mathcal{Y} = \{-1, +1\}$. Denote the set of weak classifiers

$$\mathcal{H} = \{h : \mathcal{X} \mapsto [-1, +1]\}$$

and assume $\mathcal{H}$ is a VC-subgraph. Hence, $\mathcal{D}(\mathcal{H}, \varepsilon, d_x) \le K \cdot V \log 2/\varepsilon$. A voting algorithm outputs

$$f = \sum_{i=1}^{T} \lambda_i h_i, \text{ where } h_i \in \mathcal{H}, \ \sum_{i=1}^{T} \lambda_i \le 1, \ \lambda_i > 0.$$

Let

$$\mathcal{F} = \text{conv } \mathcal{H} = \left\{ \sum_{i=1}^{T} \lambda_i h_i, \ h_i \in \mathcal{H}, \ \sum_{i=1}^{T} \lambda_i \le 1, \ \lambda_i \ge 0, \ T \ge 1 \right\}.$$

Then $\text{sign}(f(x))$ is the prediction of the label y. Let

$$\mathcal{F}_d = \text{conv}_d \ \mathcal{H} = \left\{ \sum_{i=1}^{d} \lambda_i h_i, \ h_i \in \mathcal{H}, \ \sum_{i=1}^{T} \lambda_i \le 1, \ \lambda_i \ge 0 \right\}.$$

Theorem 17.1. *For any* $x = (x_1, \ldots, x_n)$, *if*

$$\log \mathcal{D}(\mathcal{H}, \varepsilon, d_x) \le KV \log 2/\varepsilon$$

then

$$\log \mathcal{D}(conv_d \ \mathcal{H}, \varepsilon, d_x) \le KVd \log 2/\varepsilon.$$

Proof. Let $h^1, \ldots, h^D$ be ε-packing of $\mathcal{H}$ with respect to d_x, $D = \mathcal{D}(\mathcal{H}, \varepsilon, d_x)$.

Note that d_x is a norm.

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

If $f = \sum_{i=1}^{d} \lambda_i h_i$, for all h_i we can find h^{k_i} such that $d(h_i, h^{k_i}) \le \varepsilon$. Let $f' = \sum_{i=1}^{d} \lambda_i h^{k_i}$. Then

$$d(f, f') = \|f - f'\|_x = \left\| \sum_{i=1}^{d} \lambda_i (h_i - h^{k_i}) \right\|_x \le \sum_{i=1}^{d} \lambda_i \|h_i - h^{k_i}\|_x \le \varepsilon.$$

Define

$$\mathcal{F}_{D,d} = \left\{ \sum_{i=1}^{d} \lambda_i h_i, \ h_i \in \{h^1, \ldots, h^D\}, \ \sum_{i=1}^{d} \lambda_i \le 1, \ \lambda_i \ge 0 \right\}.$$

Hence, we can approximate any $f \in \mathcal{F}_d$ by $f' \in \mathcal{F}_{D,d}$ within ε.

Now, let $f = \sum_{i=1}^{d} \lambda_i h_i \in \mathcal{F}_{D,d}$ and consider the following construction. We will choose $Y_1(x), \ldots, Y_k(x)$ from $h_1, \ldots, h_d$ according to $\lambda_1, \ldots, \lambda_d$:

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

Note that with this construction

$$\mathbb{E}Y_j(x) = \sum_{i=1}^{d} \lambda_i h_i(x) = f(x).$$

Furthermore,

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{k}\sum_{j=1}^{k} Y_j - f\right\|_x^2 &= \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{k}\sum_{j=1}^{k} Y_j(x_i) - f(x_i)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{k}\sum_{j=1}^{k}(Y_j(x_i) - \mathbb{E}Y_j(x_i))\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{k^2}\sum_{j=1}^{k}\mathbb{E}(Y_j(x_i) - \mathbb{E}Y_j(x_i))^2 \\
&\leq \frac{4}{k}
\end{aligned}$$

because $|Y_j(x_i) - \mathbb{E}Y_j(x_i)| \leq 2$. Choose $k = 4/\varepsilon^2$. Then

$$\mathbb{E}\left\|\frac{1}{k}\sum_{j=1}^{k} Y_j - f\right\|_x^2 = \mathbb{E}d_x\left(\frac{1}{k}\sum_{j=1}^{k} Y_j, f\right)^2 \leq \varepsilon^2.$$

So, there exists a deterministic combination $\frac{1}{k}\sum_{j=1}^{k} Y_j$ such that $d_x(\frac{1}{k}\sum_{j=1}^{k} Y_j, f) \leq \varepsilon$. Define

$$\mathcal{F}'_{D,d} = \left\{\frac{1}{k}\sum_{j=1}^{k} Y_j : \ k = 4/\varepsilon^2, \ Y_j \in \{h_1, \ldots, h_d\} \subseteq \{h^1, \ldots, h^D\}\right\}$$

Hence, we can approximate any $f = \sum_{i=1}^{d} \lambda_i h_i \in \mathcal{F}_{D,d}$, $h_i \in \{h^1, \ldots, h^D\}$, by $f' \in \mathcal{F}'_{D,d}$ within ε.

Let us now bound the cardinality of $\mathcal{F}'_{D,d}$. To calculate the number of ways to choose k functions out of $h_1, \ldots, h_d$, assume each of h_i is chosen k_d times such that $k = k_1 + \ldots + k_d$.

We can formulate the problem as finding the number of strings of the form

$$\underbrace{00\ldots0}_{k_1}1\underbrace{00\ldots0}_{k_2}1\ldots1\underbrace{00\ldots0}_{k_d}.$$

In this string, there are $d-1$ "1"s and k "0"s, and total length is $k+d-1$. The number of such strings is $\binom{k+d-1}{k}$. Hence,

$$\begin{aligned}
\text{card } \mathcal{F}'_{D,d} &\leq \binom{D}{d} \times \binom{k+d}{k} \\
&\leq \frac{D^{D-d}D^d}{d^d(D-d)^{D-d}} \frac{(k+d)^{k+d}}{k^k d^d} \\
&= \left(\frac{D(k+d)}{d^2}\right)^d \left(\frac{D}{D-d}\right)^{D-d} \left(\frac{k+d}{k}\right)^k \\
&= \left(\frac{D(k+d)}{d^2}\right)^d \left(1+\frac{d}{D-d}\right)^{D-d} \left(1+\frac{d}{k}\right)^k \\
&\quad \text{using inequality } 1+x \leq e^x \\
&\leq \left(\frac{D(k+d)e^2}{d^2}\right)^d
\end{aligned}$$

where $k = 4/\varepsilon^2$ and $D = \mathcal{D}(\mathcal{F}, \varepsilon, d_x)$.

Therefore, we can approximate any $f \in \mathcal{F}_d$ by $f'' \in \mathcal{F}_{D,d}$ within ε and $f'' \in \mathcal{F}_{D,d}$ by $f' \in \mathcal{F}'_{D,d}$ within ε. Hence, we can approximate any $f \in \mathcal{F}_d$ by $f' \in \mathcal{F}'_{D,d}$ within 2ε. Moreover,

$$\begin{aligned}
\log \mathcal{N}(\mathcal{F}_d = \text{conv}_d\, \mathcal{H}, 2\varepsilon, d_x) &\leq d \log \frac{e^2 D(k+d)}{d^2} \\
&= d\left(2 + \log D + \log \frac{k+d}{d^2}\right) \\
&\leq d\left(2 + KV \log \frac{2}{\varepsilon} + \log\left(1+\frac{4}{\varepsilon^2}\right)\right) \\
&\leq KVd \log \frac{2}{\varepsilon}
\end{aligned}$$

since $\frac{k+d}{d^2} \leq 1+k$ and $d \geq 1$, $V \geq 1$. $\square$