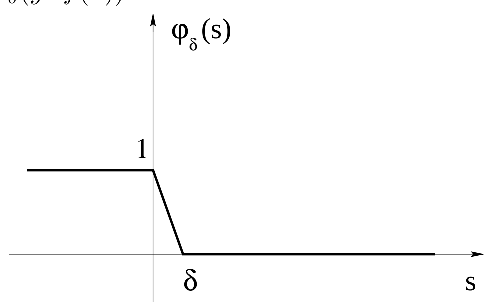
In a classification setup, we are given $\{(x_i, y_i) : x_i \in \mathcal{X}, y_i \in \{-1, +1\}\}_{i=1,\cdots,n}$, and are required to construct a classifier $y = \text{sign}(f(x))$ with minimum testing error. For any x, the term $y \cdot f(x)$ is called **margin** can be considered as the confidence of the prediction made by $\text{sign}(f(x))$. Classifiers like SVM and AdaBoost are all **maximal margin classifiers**. Maximizing margin means, penalizing small margin, controling the complexity of all possible outputs of the algorithm, or controling the generalization error.

We can define $\phi_\delta(s)$ as in the following plot, and control the error $\mathbb{P}(y \cdot f(x) \le 0)$ in terms of $\mathbb{E}\phi_\delta(y \cdot f(x))$:

$$\begin{aligned}
\mathbb{P}(y \cdot f(x) \le 0) &= \mathbb{E}_{x,y} I(y \cdot f(x) \le 0) \\
&\le \mathbb{E}_{x,y}\phi_\delta(y \cdot f(x)) \\
&= \mathbb{E}\phi_\delta(y \cdot f(x)) \\
&= \underbrace{\mathbb{E}_n\phi_\delta(y \cdot f(x))}_{\text{observed error}} + \underbrace{(\mathbb{E}(y \cdot f(x)) - \mathbb{E}_n\phi_\delta(y \cdot f(x)))}_{\text{generalization capability}},
\end{aligned}$$

where $\mathbb{E}_n\phi_\delta(y \cdot f(x)) \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^n \phi_\delta(y \cdot f(x))$.



Let us define $\phi_\delta(y\mathcal{F}) \stackrel{\triangle}{=} \{\phi_\delta(y \cdot f(x)) : f \in \mathcal{F}\}$. The function ϕ_δ satisfies Lipschetz condition $|\phi_\delta(a) - \phi_\delta(b)| \le \frac{1}{\delta}|a - b|$. Thus given any $\{z_i = (x_i, y_i)\}_{i=1,\cdots,n}$,

$$\begin{aligned}
d_z(\phi_\delta(y \cdot f(x)), \phi_\delta(y \cdot g(x))) &= \left(\frac{1}{n}\sum_{i=1}^n (\phi_\delta(y_i f(x_i)) - \phi_\delta(y_i \cdot g(x_i)))^2\right)^{1/2} &&\text{,definition of } d_z \\
&\le \frac{1}{\delta}\left(\frac{1}{n}\sum_{i=1}^n (y_i f(x_i) - y_i \cdot g(x_i))^2\right)^{1/2} &&\text{,Lipschetz condition} \\
&= \frac{1}{\delta} d_x(f(x), g(x)) \text{ ,definition of } d_x,
\end{aligned}$$

and the packing numbers for $\phi_\delta(y\mathcal{F})$ and $\mathcal{F}$ satisfies inequality $D(\phi_\delta(y\mathcal{F}), \epsilon, d_z) \le D(\mathcal{F}, \epsilon \cdot \delta, d_x)$.

Recall that for a VC-subgraph class $\mathcal{H}$, the packing number satisfies $D(\mathcal{H}, \epsilon, d_x) \le C(\frac{1}{\epsilon})^V$, where C is a constant, and V is a constant. For its corresponding VC-hull class, there exists $K(C, V)$, such that $\log D(\mathcal{F} = \text{conv}(\mathcal{H}), \epsilon, d_x) \le K(\frac{1}{\epsilon})^{\frac{2V}{V+2}}$. Thus $\log D(\phi_\delta(y\mathcal{F}), \epsilon, d_z) \le \log D(\mathcal{F}, \epsilon \cdot \delta, d_x) \le K(\frac{1}{\epsilon \cdot \delta})^{\frac{2V}{V+2}}$.

On the other hand, for a VC-subgraph class $\mathcal{H}$, $\log D(\mathcal{H}, \epsilon, d_x) \le KV \log \frac{2}{\epsilon}$, where V is the VC dimension of $\mathcal{H}$. We proved that $\log D(\mathcal{F}_d = \text{conv}_d\mathcal{H}, \epsilon, d_x) \le K \cdot V \cdot d \log \frac{2}{\epsilon}$. Thus $\log D(\phi_\delta(y\mathcal{F}_d), \epsilon, d_x) \le K \cdot V \cdot d \log \frac{2}{\epsilon\delta}$.

Remark 19.1. For a VC-subgraph class $\mathcal{H}$, let V is the VC dimension of $\mathcal{H}$. The packing number satisfies $D(\mathcal{H},\epsilon,d_x) \le \left(\frac{k}{\epsilon}\log\frac{k}{\epsilon}\right)^V$. D Haussler (1995) also proved the following two inequalities related to the packing number: $D(\mathcal{H},\epsilon,\|\cdot\|_1) \le \left(\frac{k}{\epsilon}\right)^V$, and $D(\mathcal{H},\epsilon,d_x) \le K\left(\frac{1}{\epsilon}\right)^V$.

Since conv($\mathcal{H}$) satisfies the **uniform entroy condition** (Lecture 16) and $f \in [-1,1]^{\mathcal{X}}$, with a probability of at least $1-e^{-u}$,

$$\begin{aligned}\mathbb{E}\phi_\delta(y\cdot f(x)) - \mathbb{E}_n\phi_\delta(y\cdot f(x)) &\le \frac{K}{\sqrt{n}}\int_0^{\sqrt{\mathbb{E}\phi_\delta}}\sqrt{\left(\frac{1}{\epsilon\cdot\delta}\right)^{\frac{2V}{V+2}}}d\epsilon + K\sqrt{\frac{\mathbb{E}\phi_\delta\cdot u}{n}} \\ &= Kn^{-\frac{1}{2}}\delta^{-\frac{V}{V+2}}\left(\mathbb{E}\phi_\delta\right)^{\frac{1}{V+2}} + K\sqrt{\frac{\mathbb{E}\phi_\delta\cdot u}{n}} \qquad (19.1)\end{aligned}$$

for all $f \in \mathcal{F} = \text{conv}\mathcal{H}$. The term $\mathbb{E}\phi_\delta$ to estimate appears in both sides of the above inequality. We give a bound $\mathbb{E}\phi_\delta \le x^*(\mathbb{E}_n\phi_\delta, n, \delta)$ as the following. Since

$$\begin{aligned}&\mathbb{E}\phi_\delta \le \mathbb{E}_n\phi_\delta \\ &\mathbb{E}\phi_\delta \le Kn^{-\frac{1}{2}}\delta^{-\frac{V}{V+2}}\left(\mathbb{E}\phi_\delta\right)^{\frac{1}{V+2}} \quad\Rightarrow\quad \mathbb{E}\phi_\delta \le Kn^{-\frac{1}{2}\frac{V+2}{V+1}}\delta^{-\frac{V}{V+1}} \\ &\mathbb{E}\phi_\delta \le K\sqrt{\frac{\mathbb{E}\phi_\delta\cdot u}{n}} \quad\Rightarrow\quad \mathbb{E}\phi_\delta \le K\frac{u}{n},\end{aligned}$$

It follows that with a probability of at least $1-e^{-u}$,

$$\mathbb{E}\phi_\delta \le K\cdot\left(\mathbb{E}_n\phi_\delta + n^{-\frac{1}{2}\frac{V+2}{V+1}}\delta^{-\frac{V}{V+1}} + \frac{u}{n}\right) \qquad (19.2)$$

for some constant K. We proceed to bound $\mathbb{E}\phi_\delta$ for $\delta \in \{\delta_k = \exp(-k) : k \in \mathbb{N}\}$. Let $\exp(-u_k) = \left(\frac{1}{k+1}\right)^2 e^{-u}$, it follows that $u_k = u + 2\cdot\log(k+1) = u + 2\cdot\log(\log\frac{1}{\delta_k}+1)$. Thus with a probability of at least $1-\sum_{k\in\mathbb{N}}\exp(-u_k) = 1-\sum_{k\in\mathbb{N}}\left(\frac{1}{k+1}\right)^2 e^{-u} = 1-\frac{\pi^2}{6}\cdot e^{-u} < 1-2\cdot e^{-u}$,

$$\begin{aligned}\mathbb{E}\phi_{\delta_k}(y\cdot f(x)) &\le K\cdot\left(\mathbb{E}_n\phi_{\delta_k}(y\cdot f(x)) + n^{-\frac{1}{2}\frac{V+2}{V+1}}\delta_k^{-\frac{V}{V+1}} + \frac{u_k}{n}\right) \\ &= K\cdot\left(\mathbb{E}_n\phi_{\delta_k}(y\cdot f(x)) + n^{-\frac{1}{2}\frac{V+2}{V+1}}\delta_k^{-\frac{V}{V+1}} + \frac{u+2\cdot\log(\log\frac{1}{\delta_k}+1)}{n}\right) \qquad (19.3)\end{aligned}$$

for all $f \in \mathcal{F}$ and all $\delta_k \in \{\delta_k : k \in \mathbb{N}\}$. Since $\mathbb{P}(y\cdot f(x) \le 0) = \mathbb{E}_{x,y}I(y\cdot f(x) < 0) \le \mathbb{E}_{x,y}\phi_\delta(y\cdot f(x))$, and $\mathbb{E}_n\phi_\delta(y\cdot f(x)) = \frac{1}{n}\sum_{i=1}^n\phi_\delta(y_i\cdot f(x_i)) \le \frac{1}{n}\sum_{i=1}^n I(y_i\cdot f(x_i)\le\delta) = \mathbb{P}_n(y_i\cdot f(x_i)\le\delta)$, with probability at least $1-2\cdot e^{-u}$,

$$\mathbb{P}(y\cdot f(x)) \le 0) \le K\cdot\inf_\delta\left(\mathbb{P}_n(y\cdot f(x)\le\delta) + n^{-\frac{V+2}{2(V+1)}}\delta^{-\frac{V}{V+1}} + \frac{u}{n} + \frac{2\log(\log\frac{1}{\delta}+1)}{n}\right).$$