考虑到 $\Phi_\rho(y_i h(\boldsymbol{x}_i)) \leqslant \mathbb{I}_{y_i h(\boldsymbol{x}_i) \leqslant \rho}$, 对于经验间隔损失, 有

$$\widehat{E}_\rho(h) \leqslant \frac{1}{m}\sum_{i=1}^{m} \mathbb{I}_{y_i h(\boldsymbol{x}_i) \leqslant \rho} . \tag{4.74}$$

Φ_ρ 的导数最大为 $\frac{1}{\rho}$.

由经验间隔损失 (4.72) 可知 Φ_ρ 最多是 $\frac{1}{\rho}$-Lipschitz. 引理 4.4 表明 Lipschitz 函数和假设空间 $\mathcal{H}$ 复合后的经验 Rademacher 复杂度可以基于假设空间 $\mathcal{H}$ 的经验 Rademacher 复杂度进行表示.

证明过程参阅 [Ledoux and Talagrand, 1991].

函数的复合 $f_2 \circ f_1(\boldsymbol{x}) = f_2(f_1(\boldsymbol{x}))$.

引理 4.4 若 $\Phi : \mathbb{R} \mapsto \mathbb{R}$ 为 l-Lipschitz 函数, 则对于任意实值假设空间 $\mathcal{H}$ 有下式成立:

$$\widehat{\mathfrak{R}}_D(\Phi \circ \mathcal{H}) \leqslant l\widehat{\mathfrak{R}}_D(\mathcal{H}) . \tag{4.75}$$

下面将给出基于间隔损失函数的二分类问题支持向量机的泛化误差界.

定理 4.8 令 $\mathcal{H}$ 为实值假设空间, 给定 $\rho > 0$, 对于 $0 < \delta < 1$ 和 $h \in \mathcal{H}$, 以至少 $1-\delta$ 的概率有

$$E(h) \leqslant \widehat{E}_\rho(h) + \frac{2}{\rho}\mathfrak{R}_m(\mathcal{H}) + \sqrt{\frac{\ln\frac{1}{\delta}}{2m}} , \tag{4.76}$$

$$E(h) \leqslant \widehat{E}_\rho(h) + \frac{2}{\rho}\widehat{\mathfrak{R}}_D(\mathcal{H}) + 3\sqrt{\frac{\ln\frac{2}{\delta}}{2m}} . \tag{4.77}$$

证明 构造 $\widetilde{\mathcal{H}} = \{z = (\boldsymbol{x}, y) \mapsto yh(\boldsymbol{x}) : h \in \mathcal{H}\}$, 考虑值域为 $[0,1]$ 的假设空间 $\mathcal{F} = \{\Phi_\rho \circ f : f \in \widetilde{\mathcal{H}}\}$, 根据 (4.25) 可知对于所有 $g \in \mathcal{F}$, 以至少 $1-\delta$ 的概率有

$$\mathbb{E}[g(\boldsymbol{z})] \leqslant \frac{1}{m}\sum_{i=1}^{m} g(\boldsymbol{z}_i) + 2\mathfrak{R}_m(\mathcal{F}) + \sqrt{\frac{\ln\frac{1}{\delta}}{2m}} . \tag{4.78}$$

因此, 对 $h \in \mathcal{H}$, 以至少 $1-\delta$ 的概率有

$$\mathbb{E}[\Phi_\rho(yh(\boldsymbol{x}))] \leqslant \widehat{E}_\rho(h) + 2\mathfrak{R}_m(\Phi_\rho \circ \widetilde{\mathcal{H}}) + \sqrt{\frac{\ln\frac{1}{\delta}}{2m}} . \tag{4.79}$$

因为 $\mathbb{I}_{u \leqslant 0} \leqslant \Phi_\rho(u)$ 对任意 $u \in \mathbb{R}$ 成立, 所以 $E(h) = \mathbb{E}[\mathbb{I}_{yh(\boldsymbol{x}) \leqslant 0}] \leqslant \mathbb{E}[\Phi_\rho(yh(\boldsymbol{x}))]$, 代入 (4.96) 可知

$$E(h) \leqslant \widehat{E}_\rho(h) + 2\mathfrak{R}_m(\Phi_\rho \circ \widetilde{\mathcal{H}}) + \sqrt{\frac{\ln\frac{1}{\delta}}{2m}} \tag{4.80}$$