12.3.2 不可分情形

对较为困难的学习问题, 目标概念 c 往往不存在于假设空间 $\mathcal{H}$ 中. 假定对于任何 $h \in \mathcal{H}$, $\widehat{E}(h) \neq 0$, 也就是说, $\mathcal{H}$ 中的任意一个假设都会在训练集上出现或多或少的错误. 由 Hoeffding 不等式易知:

引理 12.1 若训练集 D 包含 m 个从分布 $\mathcal{D}$ 上独立同分布采样而得的样例, $0 < \epsilon < 1$, 则对任意 $h \in \mathcal{H}$, 有

$$P\big(\widehat{E}(h) - E(h) \geqslant \epsilon\big) \leqslant \exp(-2m\epsilon^2)\ , \tag{12.15}$$

$$P\big(E(h) - \widehat{E}(h) \geqslant \epsilon\big) \leqslant \exp(-2m\epsilon^2)\ , \tag{12.16}$$

$$P\Big(\big|E(h) - \widehat{E}(h)\big| \geqslant \epsilon\Big) \leqslant 2\exp(-2m\epsilon^2)\ . \tag{12.17}$$

推论 12.1 若训练集 D 包含 m 个从分布 $\mathcal{D}$ 上独立同分布采样而得的样例, $0 < \epsilon < 1$, 则对任意 $h \in \mathcal{H}$, 式(12.18)以至少 $1-\delta$ 的概率成立:

$$\widehat{E}(h) - \sqrt{\frac{\ln(2/\delta)}{2m}} \leqslant E(h) \leqslant \widehat{E}(h) + \sqrt{\frac{\ln(2/\delta)}{2m}}\ . \tag{12.18}$$

推论 12.1表明, 样例数目 m 较大时, h 的经验误差是其泛化误差很好的近似. 对于有限假设空间 $\mathcal{H}$, 我们有

定理 12.1 若 $\mathcal{H}$ 为有限假设空间, $0 < \delta < 1$, 则对任意 $h \in \mathcal{H}$, 有

$$P\left(\big|E(h) - \widehat{E}(h)\big| \leqslant \sqrt{\frac{\ln|\mathcal{H}| + \ln(2/\delta)}{2m}}\right) \geqslant 1 - \delta\ . \tag{12.19}$$

证明 令 $h_1, h_2, \ldots, h_{|\mathcal{H}|}$ 表示假设空间 $\mathcal{H}$ 中的假设, 有

$$\begin{aligned}
&P\big(\exists h \in \mathcal{H} : \big|E(h) - \widehat{E}(h)\big| > \epsilon\big) \\
=&P\Big(\big(\big|E_{h_1} - \widehat{E}_{h_1}\big| > \epsilon\big) \vee \ldots \vee \big(\big|E_{h_{|\mathcal{H}|}} - \widehat{E}_{h_{|\mathcal{H}|}}\big| > \epsilon\big)\Big) \\
\leqslant&\sum_{h\in\mathcal{H}} P\big(\big|E(h) - \widehat{E}(h)\big| > \epsilon\big)\ ,
\end{aligned}$$

由式(12.17)可得

$$\sum_{h\in\mathcal{H}} P\big(\big|E(h) - \widehat{E}(h)\big| > \epsilon\big) \leqslant 2|\mathcal{H}|\exp(-2m\epsilon^2)\ ,$$