统计学习方法

1.6.2 泛化误差上界

对二类分类问题，当假设空间是有限个函数的集合 $\mathcal{F}=\{f_1,f_2,\cdots,f_d\}$ 时，对任意一个函数 $f\in\mathcal{F}$，至少以概率 $1-\delta$，$0<\delta<1$，以下不等式成立：

$$R(f)\le\hat{R}(f)+\varepsilon(d,N,\delta) \quad \cdots \quad \langle 1\rangle$$

其中

$$\varepsilon(d,N,\delta)=\sqrt{\frac{1}{2N}\left(\log d+\log\frac{1}{\delta}\right)} \quad \langle 2\rangle$$

$R(f)$ 是泛化误差，$\hat{R}(f)$ 是训练误差

$R(f)=E[L(Y,f(X))]$，$\hat{R}(f)=\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))$

证明：

首先给出 Hoeffding 不等式：

设 $X_1,X_2,\cdots X_N$ 是独立随机变量，且 $X_i\in[a_i,b_i]$，$i=1,2,\cdots,N$；$\bar{X}$ 是 $X_1,X_2,\cdots X_N$ 的经验均值，即 $\bar{X}=\frac{1}{N}\sum_{i=1}^{N}X_i$，则对任意 $t>0$，以下不等式成立：

$$P[\bar{X}-E(\bar{X})\ge t]\le\exp\left[-\frac{2N^2t^2}{\sum_{i=1}^{N}(b_i-a_i)^2}\right] \quad \langle 3\rangle$$

$$P[E(\bar{X})-\bar{X}\ge t]\le\exp\left[-\frac{2N^2t^2}{\sum_{i=1}^{N}(b_i-a_i)^2}\right] \quad \langle 4\rangle$$

对任意函数 $f\in\mathcal{F}$，$\hat{R}(f)$ 是 N 个独立的随机变量 $L(Y,f(X))$ 的样本均值，$R(f)$ 是随机变量 $L(Y,f(X))$ 的期望值。如果损失函数取值于区间 $[0,1]$，即对所有 i，$[a_i,b_i]=[0,1]$，那么由 Hoeffding 不等式 $\langle 4\rangle$ 可知，对 $\varepsilon>0$，以下不等式成立：

$$P[R(f)-\hat{R}(f)\ge\varepsilon]\le\exp(-2N\varepsilon^2) \quad \langle 5\rangle$$

由于$\mathcal{F}=\{f_1, f_2, \cdots, f_d\}$是一个有限集合，故

$$P(\exists f\in\mathcal{F}: R(f)-\hat{R}(f)\geq\varepsilon) = P\left(\bigcup_{f\in\mathcal{F}}\{R(f)-\hat{R}(f)\geq\varepsilon\}\right)$$

$$\leq \sum_{f\in\mathcal{F}} P(R(f)-\hat{R}(f)\geq\varepsilon)$$

$$\leq d\exp(-2N\varepsilon^2)$$

$$P(A\cup B)\leq P(A)+P(B)$$

故：

$$P(R(f)-\hat{R}(f)<\varepsilon)\geq 1-d\exp(-2N\varepsilon^2) \rightarrow P\{X\geq\varepsilon\}=1-P\{X<\varepsilon\}$$

令 $\delta = d\exp(-2N\varepsilon^2)$

则 $P(R(f)<\hat{R}(f)+\varepsilon)\geq 1-\delta$

即至少以概率$1-\delta$有$R(f)<\hat{R}(f)+\varepsilon$