欠拟合 和 过拟合

$\left\{ \begin{array}{l} \text{拟合} \\ \text{分类都存在} \end{array} \right.$

$h_\theta(x) = g(\theta^T x)$ 线性分类

$g(z) = 1\{z \geq 0\}$ $y \in \{0, 1\}$

$S: \{(x^{(i)}, y^{(i)})\}_{i=1}^{m}$ 独立同分布

$$\hat{\varepsilon}(h_\theta) = \hat{\varepsilon}_S(h_\theta) = \frac{1}{m}\sum_{i=1}^{m} 1\{h_\theta(x^{(i)}) \neq y^{(i)}\}$$

训练误差 训练集合

ERM

经验风险最小化 $\hat{\theta} = \arg\min \hat{\varepsilon}(h_\theta)$

$H = \{h_\theta : \theta \in \mathbb{R}^{n+1}\}$ 理解上从选参数

$h_\theta: X \to \{0, 1\}$ 变为选函数

$$\text{ERM}: \hat{h} = \arg\min_{h \in H} \hat{\varepsilon}(h)$$

$\wedge \to$ 去尝试

$\varepsilon(h) \to$ 真实误差

联合界引理

$A_1 \sim A_k$ 为k个事件

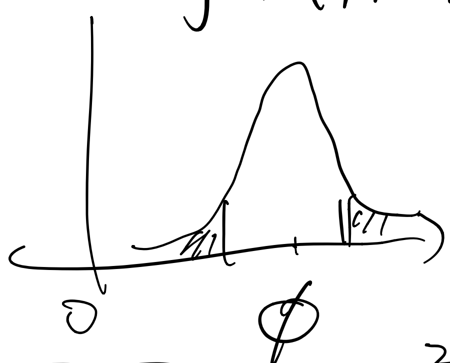
$$P(A_1 \cup A_2 \cup \cdots \cup A_k) \leq P(A_1) + P(A_2) + \cdots + P(A_k)$$

Hoeffding 不等式

$z_1, z_2 \cdots z_m$ 为 m个独立同分布 ϕ 伯努力分布

$$\hat{\phi} = \frac{1}{m}\sum_{i=1}^{m} z_i \quad 对 \forall \gamma > 0$$

$$有 P(|\hat{\phi} - \phi| > \gamma) \leq 2e^{-2\gamma^2 m}$$



$H = \{h_1, h_2, \cdots h_k\}$ k个函数

$$\hat{h} = \arg\min_{h_i \in H} \hat{\varepsilon}_S(h_i)$$

证 <1> $\hat{\varepsilon} \approx \varepsilon$

<2> $\varepsilon(\hat{h})$ 有上界

先定 $h_j \in H$

$$z_i = 1\{h_j(x^{(i)}) \neq y^{(i)}\} \in \{0, 1\}$$

$P(z_i = 1) = \varepsilon(h_j)$ $\quad z_i$'s are IID

独立同分布

$$\hat{\varepsilon}(h_j) = \frac{1}{m}\sum_{i=1}^{m} Z_i = \frac{1}{m}\sum_{i=1}^{m} 1\{h_j(x^{(i)}) = y^{(i)}\}$$

由 Hoeffding不等式

$$P(|\varepsilon(h_j) - \hat{\varepsilon}(h_j)| > \gamma) \le 2e^{-2\gamma^2 m}$$

m越大

概率上越接近

A_j为 $|\varepsilon(h_j) - \hat{\varepsilon}(h_j)| > \gamma$ 的事件

$$P(A_j) \le 2e^{-2\gamma^2 m}$$

$$P(\exists j \in H \quad |\varepsilon(h_j) - \hat{\varepsilon}(h_j)| > \gamma)$$

$$= P(A_1 \cup A_2 \cup \cdots \cup A_k) \le \sum_{i=1}^{k} P(A_i)$$

$$\le \sum_{i=1}^{k} 2e^{-2\gamma^2 m} = 2ke^{-2\gamma^2 m}$$

1 − 两边

$$P(\text{not } \exists h_j \in H \quad |\varepsilon(h_j) - \hat{\varepsilon}(h_j)| > \gamma)$$

$$= P(\forall h_j \in H \quad |\varepsilon(h_j) - \hat{\varepsilon}(h_j)| \le \gamma) \ge 1 - 2ke^{-2\gamma^2 m}$$

一致收敛的概率

(对无限H不适用，概率上趋近)

给定γ和允许错误率δ需要多大的训练集合

$$m \ge \frac{1}{2\gamma^2}\log\frac{2k}{\delta}$$

~~计算机中 $\forall k,\ \log k \le 30$~~

设 $\forall h \in H$ $|\varepsilon(h) - \hat{\varepsilon}(h)| \le \gamma$

最小训练误差 $\hat{h} = \arg\min_{h \in H} \hat{\varepsilon}(h)$

$$h^* = \arg\min_{h \in H} \varepsilon(h)$$

$$\begin{aligned}\varepsilon(\hat{h}) &\le \hat{\varepsilon}(\hat{h}) + \gamma \\ &\le \hat{\varepsilon}(h^*) + \gamma \\ &\le \varepsilon(h^*) + 2\gamma\end{aligned}$$

上面联系起来

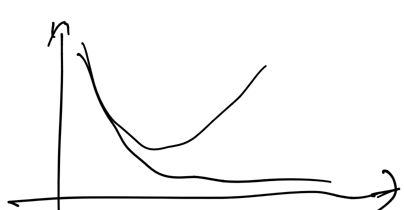
令 $|H| = k$ 有限，m 和 δ 固定，要 $1-\delta$ 的准确度

$$\varepsilon(\hat{h}) \le \Big(\underbrace{\min_{h \in H} \varepsilon(h)}_{\varepsilon(h^*)}\Big) + 2\sqrt{\frac{1}{2m}\log\frac{2k}{\delta}}$$

数据误差　　假设误差

令 $\underbrace{\gamma = \sqrt{\frac{1}{2m}\log\frac{2k}{\delta}}}$ 也有

下成立



[有限维H]

这一章主要是定量的分析模型误差