机器学习基础作业 7

2025 年 5 月 17 日

问题 1. 设 $\mathcal{X}$ 为一个域, $D_1, D_2, \ldots, D_m$ 是 $\mathcal{X}$ 上的一系列分布. 设 $\mathcal{H}$ 是 $\mathcal{X}$ 上的一个有限二元分类器类, 且 $f \in \mathcal{H}$. 假设我们得到一个样本 S, 包含 m 个独立但不同分布的实例; 第 i 个实例从 D_i 中采样, 然后 y_i 被设为 $f(x_i)$. 定义 $\bar{D}_m$ 为平均值, 即

$$\bar{D}_m = (D_1 + \cdots + D_m)/m.$$

固定一个精度参数 $\epsilon \in (0, 1)$. 证明:

$$P\left[\exists h \in \mathcal{H} : L_{(\bar{D}_m, f)}(h) > \epsilon, L_{(S,f)}(h) = 0\right] \le |\mathcal{H}|e^{-\epsilon m}$$

提示: 使用算术-几何平均不等式.

证明. $L(S, f)(h) = 0$ 意味着 $h(x_i) = y_i$ 对所有 $i \in [m]$ 成立. 设 $p_i = P_{D_i}(h(x_i) \neq f(x_i))$, 则:

$$L_{(\bar{D}_m, f)}(h) = \frac{1}{m}\sum_{i=1}^{m} p_i > \epsilon$$

因而, 在 $L_{(\bar{D}_m, f)}(h) > \epsilon$ 的前提下, 可得:

$$P\left[L_{(S,f)}(h) = 0\right] = \prod_{i \in [m]} P_{D_i}(h(x_i) = f(x_i)) = \prod_{i \in [m]} (1 - p_i)$$

根据算术-几何平均不等式, 我们有:

$$\prod_{i \in [m]} (1 - p_i) \le \left(1 - \frac{1}{m}\sum_{i=1}^{m} p_i\right)^m < (1 - \epsilon)^m < e^{-\epsilon m}$$

从而:

$$P\left[\exists h \in \mathcal{H} : L_{(\bar{D}_m, f)}(h) > \epsilon, L_{(S,f)}(h) = 0\right] \le |\mathcal{H}|e^{-\epsilon m}$$

□

问题 2. 设 $\mathcal{H}$ 是一个二元分类器的假设类. 证明如果 $\mathcal{H}$ 是不可知 *PAC* 可学习的, 那么 $\mathcal{H}$ 也是 *PAC* 可学习的. 此外, 如果 A 是 $\mathcal{H}$ 的一个成功的不可知 *PAC* 学习器, 那么 A 也是 $\mathcal{H}$ 的一个成功的 *PAC* 学习器.

证明. 由于 $\mathcal{H}$ 是不可知 PAC 可学习的, 所以存在一个算法 A 和一个常数 m_0, 使得对于所有的 $\epsilon > 0$, $\delta > 0$, 当 $m \ge m_0$ 时, 有:

$$P\left[L_{(S,f)}(h) \le \frac{\epsilon}{2} + \min_{h' \in \mathcal{H}} L_{(S,f)}(h')\right] \ge 1 - \delta$$

而 $\mathcal{H}$ 是二分类器的假设类, 则存在一个假设 $h' \in \mathcal{H}$ 使得 $L_{(S,f)}(h') < \frac{\epsilon}{2}$. 于是我们有:

$$P\left[L_{(S,f)}(h) \le \epsilon\right] \ge 1 - \delta$$

□

问题 3. *给出如下概念的定义:*

- *PAC 可学习性*
- *不可知 PAC 可学习性*
- *一致收敛性质, 并解释其中 "一致" 的含义*
- *估计误差*

证明. • PAC 可学习性: 设 $\mathcal{H}$ 是一个假设类, D 是一个分布, S 是从 D 中采样的样本. 如果存在一个算法 A 和一个常数 m_0, 使得对于所有的 $\epsilon > 0$, $\delta > 0$, 当 $m \geq m_0$ 时, 有:

$$P\left[L_{(S,f)}(h) \leq \epsilon\right] \geq 1 - \delta$$

对于所有的 $h \in \mathcal{H}$ 和所有的 $f \in \mathcal{H}$, 则称 $\mathcal{H}$ 是 PAC 可学习的.

- 不可知 PAC 可学习性: 如果 $\mathcal{H}$ 是不可知 PAC 可学习的, 则存在一个算法 A 和一个常数 m_0, 使得对于所有的 $\epsilon > 0$, $\delta > 0$, 当 $m \geq m_0$ 时, 有:

$$P\left[L_{(S,f)}(h) \leq \epsilon + \min_{h' \in \mathcal{H}} L_{(S,f)}(h')\right] \geq 1 - \delta$$

对于所有的 $h \in \mathcal{H}$ 和所有的 $f \in D$, 则称 $\mathcal{H}$ 是不可知 PAC 可学习的.

- 一致收敛性质: 一致收敛性质是指对于所有的假设类 $\mathcal{H}$ 和所有的分布 D, 存在一个常数 m_0, 使得当样本大小 m 大于等于 m_0 时, 对于所有的假设类中的假设 h, 有:

$$|L_{(S,f)}(h) - L_{(D,f)}(h)| < \epsilon$$

"一致" 的含义是对于所有的假设类中的假设都成立.

- 估计误差: 估计误差是指在有限样本上理论最小可达误差和最小经验误差之间的差值.

□

问题 4. *令 A 为满足如下条件的学习算法: 如果 $m \geq m_{\mathcal{H}}(\epsilon)$ 则对任一分布 D 有:*

$$\mathbb{E}_{S \sim D^m}[L_D(A(S))] \leq \min_{h \in \mathcal{H}} L_D(h) + \epsilon$$

这里 $A(S)$ 为学习算法 A 基于训练数据集 S 从假设空间 $\mathcal{H}$ 中选择的假设. 证明: 对任一 $\delta \in (0,1)$, 如果 $m \geq m_{\mathcal{H}}(\epsilon\delta)$, 则以至少 $1-\delta$ 的概率有:

$$L_D(A(S)) \leq \min_{h \in \mathcal{H}} L_D(h) + \epsilon$$

证明. 我们设:

$$Y = L_D(A(S)) - \min_{h \in \mathcal{H}} L_D(h)$$

则由题目条件可得 $E(Y) \leq \epsilon\delta$. 由 Markov 不等式可得:

$$P(Y \geq \epsilon) \leq \frac{E(Y)}{\epsilon} \leq \frac{\epsilon\delta}{\epsilon} = \delta$$

即得到至少有 $1-\delta$ 的概率有:

$$L_D(A(S)) \leq \min_{h \in \mathcal{H}} L_D(h) + \epsilon$$

□