机器学习基础作业 9

2025 年 6 月 4 日

问题 1. 简述非一致可学习性和不可知 *PAC* 可学习性之间的关系.

证明. 两个算法都强调了泛化能力. 即只要训练样本 m 充分大, 就可以保证算法以充分大的概率让损失函数比最优损失函数相差充分小.

区别在于, 非一致可学习性的样本集大小 m 与每个假设 h 是相关的. 换言之不可知 PAC 可学习性一定是非一致可学习性的. □

问题 2. 设 $\mathcal{H}_1$ 和 $\mathcal{H}_2$ 为假设空间, 令 $\mathcal{H} = \{h_1 \cdot h_2 \mid h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\}$, 证明增长函数满足

$$\Pi_{\mathcal{H}}(m) \leq \Pi_{\mathcal{H}_1}(m)\Pi_{\mathcal{H}_2}(m)$$

证明. 考虑 C 满足 $|C| = m$, 且使得 $|\mathcal{H}_C| = \Pi_{\mathcal{H}}(m)$.

构建如下映射 $f : \mathcal{H}_1 \times \mathcal{H}_2 \mapsto \mathcal{H}_C$:

$$(h_1(c_1), h_1(c_2), \ldots, h_1(c_m)) \in \mathcal{H}_1, (h_2(c_1), h_2(c_2), \ldots, h_2(c_m)) \in \mathcal{H}_2 \to (h(c_1), h(c_2), \ldots, h(c_m)) \in \mathcal{H}_C$$

其中 $h = h_1 \cdot h_2$, 因此可知 f 是良定的, 且是满射的. 因此 $\Pi_{\mathcal{H}}(m) \leq |\mathcal{H}_{1C}| \cdot |\mathcal{H}_{2C}| = \Pi_{\mathcal{H}_1}(m) \cdot \Pi_{\mathcal{H}_2}(m)$. □

问题 3. 若 $\mathcal{H}_1$ 和 $\mathcal{H}_2$ 都是 $\mathcal{X}$ 到 $\{0, 1\}$ 的函数集, 且二者的 *VC* 维有限, 证明:

$$VCdim(\mathcal{H}_1 \cup \mathcal{H}_2) \leq VCdim(\mathcal{H}_1) + VCdim(\mathcal{H}_2) + 1$$

证明. 设 $\mathrm{VCdim}(\mathcal{H}) = m, \mathrm{VCdim}(\mathcal{H}_1) = m_1, \mathrm{VCdim}(\mathcal{H}_2) = m_2$.

对于可被 $\mathcal{H}_1 \cup \mathcal{H}_2$ 打散的集合 $C = \{c_1, c_2, \ldots, c_m\}$, 考虑 $\mathcal{H}_1$ 和 $\mathcal{H}_2$ 在 C 上的限制. 由 Sauer-Shelah-Perles 引理可知:

$$|\mathcal{H}_{1C}| \leq \sum_{k=0}^{m_1} \binom{m}{k}, \quad |\mathcal{H}_{2C}| \leq \sum_{k=0}^{m_2} \binom{m}{k}$$

$$2^m \leq |\mathcal{H}_C| \leq |\mathcal{H}_{1C}| + |\mathcal{H}_{2C}| \leq \sum_{k=0}^{m_1} \binom{m}{k} + \sum_{k=0}^{m_2} \binom{m}{k}$$

如果 $m > m_1 + m_2 + 1$, 则:

$$2^m = \sum_{k=0}^{m} \binom{m}{k} > \sum_{k=0}^{m_1} \binom{m}{k} + \sum_{k=m-m_2}^{m} \binom{m}{k} = \sum_{k=0}^{m_1} \binom{m}{k} + \sum_{k=0}^{m_2} \binom{m}{k}$$

由此和上式矛盾, 因此 $m \leq m_1 + m_2 + 1$.

□