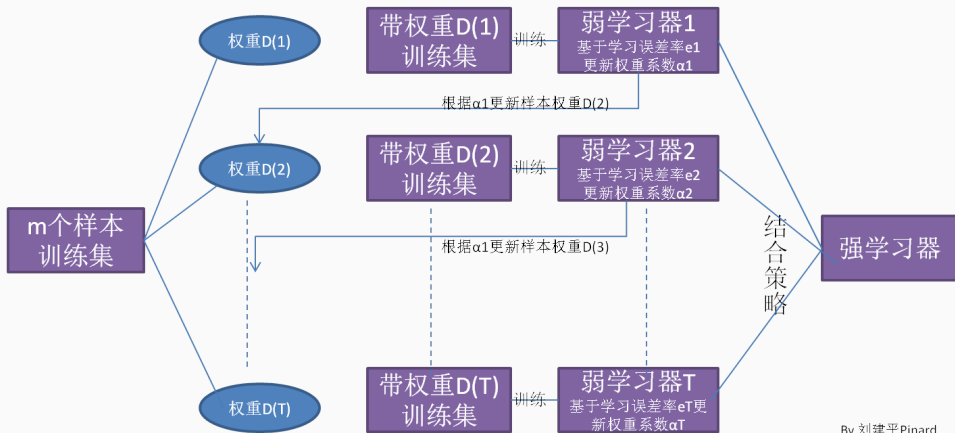

分类问题视角下的 AdaBoost 算法

November 23, 2025

AdaBoost 模型结构

- AdaBoost 是一种典型的提升（Boosting）型集成学习框架，其核心思想是：**通过多轮训练、聚焦难样本，把一群“弱学习器”提升为一个“强学习器”。**
- 在每一轮中，AdaBoost 都会为总模型增加一个新的学习器，指导模型的弱学习器个数达到预先指定的值。
- 训练新学习器时，根据上一轮的推理结果在同一训练集上**重新分配样本权重**，使新的弱学习器更加关注上一轮中被分错或“难学”的样本。
- 各轮得到的弱学习器本身能力都比较弱，但在最后通过加权组合（加权投票或加权求和），形成一个整体性能更高、泛化能力更强的强学习器。
- AdaBoost 不限定弱学习器的具体形式（如决策树桩、小深度决策树等），因此可以看作一个**通用的、可移植的集成学习框架。**

AdaBoost 模型结构图



By 刘建平Pinard

Figure 1. AdaBoost 模型结构图

AdaBoost 模型结构

计算流程

错误率计算规则

假设训练集包含 m 个样本 $\{(x_i, y_i), \dots, (x_m, y_m)\}$ 。

不失一般性地，假设现在训练已经进行到第 t 轮，将要训练第 t 个弱学习器。

记第 t 个学习器在样本 x 上的预测结果为 $h_t(x)$ 。

设训练第 t 个弱学习器所使用的样本权重为

$$D_t = \{\omega_{t1}, \dots, \omega_{tm}\}, \quad \sum_{i=1}^m \omega_{ti} = 1.$$

则第 t 轮弱学习器的加权错误率为

$$\varepsilon_t = \sum_{i=1}^m \omega_{ti} \mathbf{1}\{h_t(x_i) \neq y_i\}.$$

样本加权规则

第 t 轮 ($t = 1, \dots, T$):

1. 计算系数

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}, \quad 0 < \beta_t < 1.$$

2. 先得到更新后的未归一化权重

$$\tilde{\omega}_{t+1,i} = \omega_{ti} \beta_t^{1 - \mathbf{1}\{h_t(x_i) \neq y_i\}}, \quad i = 1, \dots, m.$$

3. 再将其归一化, 得到下一轮的权重分布

$$\omega_{t+1,i} = \frac{\tilde{\omega}_{t+1,i}}{\sum_{j=1}^m \tilde{\omega}_{t+1,j}}.$$

直观理解:

$$\begin{cases} h_t(x_i) = y_i \Rightarrow \tilde{\omega}_{t+1,i} = \omega_{ti} \beta_t & (\text{分对: 权重减小}) \\ h_t(x_i) \neq y_i \Rightarrow \tilde{\omega}_{t+1,i} = \omega_{ti} & (\text{分错: 权重不变, 归一化后相对增大}) \end{cases}$$

收敛性

设:

- 第 t 轮弱学习器的加权错误率为 ε_t ;
- 最终强分类器为 h_f , 其在训练分布 D 下的错误率为

$$\varepsilon = \Pr_{i \sim D}[h_f(x_i) \neq y_i].$$

训练误差上界

$$\varepsilon \leq 2^T \prod_{t=1}^T \sqrt{\varepsilon_t(1 - \varepsilon_t)}.$$

引入第 t 轮的“优势” (edge) : 比随机猜测强的部分

$$\gamma_t = \frac{1}{2} - \varepsilon_t,$$

则上界可以改写为

$$\varepsilon \leq \prod_{t=1}^T \sqrt{1 - 4\gamma_t^2} = \exp\left(-\sum_{t=1}^T \text{KL}\left(\frac{1}{2} \parallel \frac{1}{2} - \gamma_t\right)\right) \leq \exp\left(-2\sum_{t=1}^T \gamma_t^2\right).$$

特殊情形：若所有弱学习器的错误率都相同， $\varepsilon_t = \frac{1}{2} - \gamma$ ($\gamma > 0$)，则

$$\varepsilon \leq (1 - 4\gamma^2)^{T/2} = \exp(-T \cdot \text{KL}(\frac{1}{2} \parallel \frac{1}{2} - \gamma)) \leq \exp(-2T\gamma^2).$$

结论：只要每一轮的弱学习器都略好于随机猜测 ($\gamma_t > 0$)，AdaBoost 在训练集上的错误率会随轮数 T **指数级下降**。

收敛性

收敛速度证明 (可选)

任务：手写数字识别

数据集

关键代码

准确率

任务：手写数字识别

误差分析

针对 AdaBoost 对噪声（离群点）敏感的问题，从过拟合的角度入手分析

AdaBoost 会逐渐给离群样本赋予更高的权重，追踪样本权重的变化

MINIST 数据集貌似是清洗过的，我们可以自行添加噪声来呈现 AdaBoost 的过拟合效应

谢谢大家！