Boosting 减方偏差

弱可学习算法 ⟶ 强学习算法

Adaboost

- 整体数据加：提升该分类器权重
- 单个样本加：降低该样本权重

算法流程

1. 初始化权值分布（均匀分布）

$$D_1 = (w_{11}, \cdots w_{1i}, \cdots w_{1N}) \quad w_{1i} = \frac{1}{N}$$

2. 对 $m = 1, \cdots m$

(1) 使用权值分布 D_m 训练得到基本分类器 G_m

(2) 计算 G_m 在训练数据上的误差率 $e_m = \sum_{i=1}^{N} I(G_m(x_i) \neq y_i) w_{mi}$

(3) 计算 $G_m(x)$ 在训练数据集上的分类误差 $\alpha_m = \frac{1}{2} \log \frac{1-e_m}{e_m}$

(4) 更新样本权值分布 $D_{m+1} = (w_{m+1,1} \cdots w_{m+1,N})$

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(x_i)) \begin{cases} \frac{w_{mi}}{Z_m} e^{-\alpha_m} & G_i = y_i \\ \frac{w_{mi}}{Z_m} e^{\alpha_m} & G_i \neq y_i \end{cases}$$

Z_m：规范因子 $= \sum_{i=1}^{N} w_{mi} \exp(-\alpha_m y_i G_m(x_i))$

本轮样本权重之和（权值之和=1）

3. 构建基本分类器线性组合

$$f(x) = \sum_{m=1}^{M} \alpha_m G_m(x) \longrightarrow G(x) = \text{sign}(f(x))$$

Bagging → 减少方差

$$\sigma^2 = \rho\sigma^2 + (1-\rho)\frac{\sigma^2}{n}$$

相关系数

提升树 Boosting tree

- 基分类器：分类/回归树
- 提升树模型

$$f_M(x) = \sum_{m=1}^{M} T(x; \Theta_m)$$

- 前向分步算法 —— 消除残差

$$f_m(x) = f_{m-1}(x) + T(x; \Theta_m)$$

$$\hat{\Theta}_m = \arg\min \sum_{i=1}^{N} L(y_i, \underbrace{f_{m-1}(x_i)}_{\text{固定}} + T(x_i; \Theta_m))$$

$$f_0(x) = 0$$

$$f_1(x) = f_0(x) + T(x; \Theta_1)$$

$$\frac{1}{N} \sum_{i=1}^{N} L(y_i, f_1(x_i))$$

对于回归问题：

$$L(y_i, f_1(x_i)) = \frac{1}{2}(y - f(x))^2$$

梯度提升算法

输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\dots,(x_N,y_N)\}$

损失函数 $L(y,f(x))$

输出：回归树 $\hat{f}(x)$

(1) 初始化

$$f_0(x)=\arg\min_c \sum_{i=1}^{N} L(y_i,c)$$

(2) 对于 $m=1,\dots,M$

(a) 对 $i=1,\dots,N$ 计算

$$r_{mi}=-\left[\frac{\partial L(y_i,f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}$$

残差近似（对于 MAE）

$$-\frac{\partial \frac{1}{2}[y_i-f(x_i)]^2}{\partial f(x_i)}=[y_i-f(x_i)]$$ 残差

(b) 对 r_{mi} 拟合一个回归树，得到第 m 棵树的叶节点区域 R_{mj} $j=1,\dots,J$

(c) 对 $j=1,\dots,J$ 计算 $C_{mj}=\arg\min_c \sum_{x\in R_{mj}} L(y_i,f_{m-1}(x_i)+c)$

(d) 更新 $f_m(x)=f_{m-1}(x)+\sum_{j=1}^{J} C_{mj} I(x\in R_{mj})$

(3) 得到回归树

$$\hat{f}(x)=f_M(x)=\sum_{m=1}^{M}\sum_{j=1}^{J} C_{mj} I(x\in R_{mj})$$

Adaboost

前向分步算法

$$f_m(x)=f_{m-1}(x)+\alpha_m G_m(x)$$

$$(\alpha_m,G_m(x))$$

$$=\arg\min_{\alpha,G}\sum_{i=1}^{N}\exp[-y_i(f_{m-1}(x_i)+\alpha_m G_m(x_i))]$$

$$=\arg\min_{\alpha,G}\sum_{i=1}^{N}\overline{w}_{mi}\exp[-y_i\alpha_m G_m(x_i)]$$

$$=\arg\min_{\alpha,G}\sum_{\text{正确}}\overline{w}_{mi}\exp[-\alpha_m]+\arg\min_{\alpha,G}\sum_{\text{错误}}\overline{w}_{mi}\exp[\alpha_m]$$

$$=\arg\min_{\alpha,G}\sum_{\text{正}}\overline{w}_{mi}\exp[-\alpha_m]+\arg\min_{\alpha,G}\sum_{\text{错}}\overline{w}_{mi}\exp[-\alpha_m]-\arg\min_{\alpha,G}\sum_{\text{错}}\overline{w}_{mi}\exp[-\alpha_m]+\arg\min_{\alpha,G}\sum_{\text{错}}\overline{w}_{mi}\exp[\alpha_m]$$

$$=\exp[-\alpha_m]\sum_{i=1}^{N}\overline{w}_{mi}+\left(e^{\alpha_m}-e^{-\alpha_m}\right)\sum\overline{w}_{mi}I(y_i\neq G(x_i)) \quad ①$$

G 的最优解

$$G_m^*=\arg\min_G\sum_i\overline{w}_{mi}I(y_i\neq G(x_i))$$

α 的最优解

对 ① 求导，令 $\frac{\partial①}{\partial\alpha}=0$

$$-e^{-\alpha}\sum_{i=1}^{N}\overline{w}_{mi}+(e^{\alpha}+e^{-\alpha})\sum\overline{w}_{mi}I(y_i\neq G(x_i))=0$$

$$\frac{e^{\alpha}+e^{-\alpha}}{e^{-\alpha}}=\frac{1/\sum\overline{w}_{mi}I(y_i\neq G(x_i))}{1/\sum\overline{w}_{mi}}=\frac{1}{e_m}$$

$$e^{\alpha}=\frac{1}{e_m}-1$$

$$\alpha=\frac{1}{2}\log\frac{1-e_m}{e_m}$$

权重更新

$$\overline{w}_{mi}=\exp(-y_i f_{m-1}(x_i))$$

$$=\exp\left(-y_i\sum_{j=1}^{m-1}\alpha_j G_j(x_i)\right)$$

$$=\prod_j\exp(-y_i\alpha_j G_j(x_i))$$

$$\therefore \overline{w}_{m+1,i}=\overline{w}_{mi}\cdot\exp(-y_i\alpha_m G_m(x_i))$$