

提升方法

W.J.Z

0.1 AdaBoost 算法

对提升方法而言，有两个问题需要回答：(1): 在每一轮中如何改变训练数据的权值或概率分布;(2): 如何将弱分类器组合成强分类器。AdaBoost 针对第一个问题：提高被前一轮弱分类器错误分类样本的权值，降低被正确分类样本的权值; 针对第二个问题：加大分类误差率小的弱分类器的权值，减小分类误差率大的弱分类器的权值。

Algorithm 1: AdaBoost

Input: 训练数据集 T , 弱分类算法

Output: 最终分类器

1 初始化训练数据的权值分布

$$D_1 = (w_{11}, \dots, w_{1i}, \dots, w_{1N}), w_{1i} = \frac{1}{N}, i = 1, 2, \dots, N$$

2 对 $m = 1, 2, \dots, M$: (1): 对具有权值分布的 D_m 的训练数据集学习，得到基本分类器 $G_m(x)$ 。(2): 计算基本分类器在训练数据集的分类误差率 $e_m = \sum_{i=1}^N w_{mi} I(G_m(x_i) \neq y_i)$;(3): 计算 $G_m(x)$ 的系数 $\alpha_m = \frac{1}{2} \log \frac{1-e_m}{e_m}$; 这里的对数是自然对数。(4): 更新训练数据集的权值分布

$$D_{m+1} = (w_{m+1,1}, \dots, w_{m+1,i}, \dots, w_{m+1,N})$$

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(x_i)), i = 1, 2, \dots, N$$

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

3 构建基本分类器的线性组合

$$f(x) = \sum_{i=1}^M \alpha_m G_m(x)$$

4 得到最终分类器:

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

0.2 提升树

对于二分类问题，提升树只需要将 AdaBoost 算法的基本分类器限制为二类分类树即可。对于回归提升树：

Algorithm 2: 回归问题的提升树算法

Input: 训练数据集

Output: 提升树 $f_M(x)$

1 初始化 $f_0(x)$ 对 $m = 1, 2, \dots, M$ 计算残差

$$r_{mi} = y_i - f_{m-1}(x_i) \quad i = 1, 2, \dots, N$$

2 拟合残差 r_{mi} 学习一个回归树，得到 $T(x; \Theta_m)$

3 更新

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

4 得到回归问题提升树

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