

AdaBoost证明过程

给定训练集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其中 $x_i \in X \subseteq \mathbb{R}^n$, $y_i \in Y = \{-1, +1\}$,

AdaBoost训练算法如下。

(1)、初始化训练数据样本的权重分布, 即为每个训练样本分配一个初始权重:

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

(2)、对于弱分类器 $t=1, 2, \dots, T$, 分别执行以下步骤。

(a)、对包含权重分布 D_t 的训练集进行训练并得到弱分类器 $G_t(x)$ 。

(b)、计算 $G_t(x)$ 在当前加权训练集上的分类误差率 ε_t :

$$\varepsilon_t = P(G_t(x_i) \neq y_i) = \sum_{i=1}^N w_{t,i} I(G_t(x_i) \neq y_i)$$

(c)、根据分类误差率 ε_t 计算当前弱分类器的权重系数 α_t :

$$\alpha_t = \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t} \quad \dots \text{式(1)}$$

(d)、调整训练集的权重分布:

$$D_{t+1} = (w_{t+1,1}, \dots, w_{t+1,i}, \dots, w_{t+1,N}) \quad w_{t+1,i} = \frac{w_{t,i}}{Z_t} \exp(-\alpha_t \cdot y_i \cdot G_t(x_i)) \quad \dots \text{式(2)}$$

其中 Z_t 为归一化因子, $Z_t = \sum_{i=1}^N w_{t,i} \exp(-\alpha_t \cdot y_i \cdot G_t(x_i))$

(3)、最后构建 T 个弱分类器的线性组合:

$$f(x) = \sum_{t=1}^T \alpha_t G_t(x)$$

最终的强分类器写为:

$$G(x) = \text{sign}(f(x)) = \text{sign}\left(\sum_{t=1}^T \alpha_t \cdot G_t(x)\right)$$

在式(1)弱分类器系数的计算过程中, 当弱分类器的分类误差率 $\varepsilon_t \leq \frac{1}{2}$ 时, $\alpha_t \geq 0$, 且 α_t 随 ε_t 的减小而变大, 这也正是弱分类器权重计算公式的设计思想, 它能够使分类误差率较低的分类器有较大的权重系数。

式(2)训练样本权重分布可以写为:

$$w_{t+1,i} = \begin{cases} \frac{w_{t,i}}{Z_t} e^{-\alpha_t}, & G_t(x_i) = y_i \\ \frac{w_{t,i}}{Z_t} e^{\alpha_t}, & G_t(x_i) \neq y_i \end{cases}$$

当样本被弱分类器正确分类时它的权重变小; 当样本被弱分类器错误分类时, 它的权重变大。

AdaBoost 与前向分布算法

AdaBoost 以加性模型为模型、指数函数为损失函数、前向分步为算法的分类学习算法。

$$f(x) = \sum_{t=1}^T \alpha_t b(x; r_t) \quad \dots \text{式(3)}$$

其中 $b(x; r_t)$ 为基模型, r_t 为基模型参数, α_t 为基模型系数。

加性模型的目标函数为如下最小化损失函数:

$$\min_{\alpha_t, r_t} \sum_{i=1}^N L(y_i, \sum_{t=1}^T \alpha_t b(x_i; r_t))$$

针对加性模型的特点, 从前往后只优化一个基模型的参数, 每步优化叠加后便可逐步逼近上式的目标函数。每步优化的表达式如下:

$$\min_{\alpha, r} \sum_{i=1}^N L(y_i, \alpha b(x_i; r))$$

利用前向分布算法求解式(3)的优化问题过程如下:

(1) 初始化模型 $f_0(x) = 0$ 。

(2) 对于基分类器 $t = 1, 2, \dots, T$, 分别执行以下操作:

(a)、以 α_t 和 r_t 为优化参数, 最小化目标损失:

$$(\alpha_t, r_t) = \arg \min_{\alpha, r} \sum_{i=1}^N L(y_i, f_{t-1}(x_i) + \alpha b(x_i; r))$$

(b)、更新加性模型:

$$f_t(x) = f_{t-1}(x) + \alpha_t b(x; r_t)$$

(c)、可得到最后的加性模型为:

$$f(x) = f_T(x) = \sum_{t=1}^T \alpha_t b(x; r_t)$$

第 t 轮迭代得到第 t 个基分类器的权重系数 α_t 、第 t 个基模型 $G_t(x)$ 和 t 轮迭代后的加性模型 $f_t(x)$:

$$f_t(x) = f_{t-1}(x) + \alpha_t G_t(x)$$

优化目标是使 $f_t(x)$ 在给定训练集上指数损失最小化:

$$(\alpha_t, G_t(x)) = \arg \min_{\alpha, G} \sum_{i=1}^N \exp(-y_i (f_{t-1}(x_i) + \alpha G(x_i)))$$

最小化指数损失即可得到 AdaBoost 优化参数。