Adaboost

Input

- Training data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$
- Algorithm parameter: Number M of weak learners

Training algorithm

1. Initialize the observation weights $w_i = \frac{1}{n}$ for $i = 1, 2, ..., n$.
2. For $m = 1$ to M:
   2.1 Fit a classifier $g_m(x)$ to the training data using weights w_i.
   2.2 Compute
   $$\text{err}_m := \frac{\sum_{i=1}^{n} w_i \mathbb{I}\{y_i \neq g_m(x_i)\}}{\sum_i w_i}$$
   2.3 Compute $\alpha_m = \frac{1}{2}\log(\frac{1-\text{err}_m}{\text{err}_m})$
   2.4 Set $w_i \leftarrow w_i \cdot \exp(\alpha_m \cdot \mathbb{I}(y_i \neq g_m(x_i)))$ for $i = 1, 2, ..., n$.
3. Output
$$f(x) := \text{sign}\left(\sum_{m=1}^{M} \alpha_m g_m(x)\right)$$

5 / 29

adaboost = 可加模型 + 指数损失 + 基学习器二分类

可加模型

根据可加模型的定义，我们令 $F(x) = \Sigma_{m=1}^{M} \alpha_m g_m(x)$，则可知

$f(x) = sign(\Sigma_{m=1}^{M})\alpha_m g_m(x)) = sign(F_M(x))$

由此可分解 $F_M(x) = \Sigma_{m=1}^{M} \alpha_m g_m(x) = \Sigma_{m=1}^{M-1} \alpha_m g_m(x) + \alpha_M g_M(x) = F_{M-1}(x) + \alpha_M g_M(x)$.

指数损失

$L(y, f(x)) = e^{-yf(x)}$

算法推导

Adaboost 是通过前向分步学习算法一步步而得到的，即第M次迭代得到的结果取决于前M-1次的迭代加权和，因此在推导过程中，我们假设前M-1次的迭代结果已知，来针对第M次迭代的参数进行推导。

首先根据指数损失函数和可加模型的定义，可以得到损失函数：

$$\begin{aligned} Loss &= \Sigma_{i=1}^{n} e^{-y_i F(x_i)} \\ &= \Sigma_{i=1}^{n} e^{-y_i F_{M-1}(x_i) - y_i \alpha_M g_M(x_i)} \\ &= \Sigma_{i=1}^{n} e^{-y_i F_{M-1}(x_i)} e^{-y_i \alpha_M g_M(x_i)} \end{aligned}$$

算法的目标是为了求参数α_M 和 $g_M(x)$ 而使得损失函数最小，而$e^{-y_iF_{M-1}(x_i)}$ 与这两个参数无关，只与样本有关，因此令$w_{Mi} = e^{-y_iF_{M-1}(x_i)}$，从而

$$Loss = \Sigma_{i=1}^n w_{Mi} e^{-y_i \alpha_M g_M(x_i)}$$

而且因为问题为二分类问题，因此我们可知

$$-y_i g(x_i) = 2I(y_i \neq g(x_i)) - 1$$

所以

$$\begin{aligned} Loss &= \Sigma_{i=1}^n w_{Mi} e^{-y_i \alpha_M g_M(x_i)} \\ &= \Sigma_{i=1}^n w_{Mi} e^{\alpha_M [2I(y_i \neq g_M(x_i)) - 1]} \end{aligned}$$

为了得到损失函数的最小值，我们可以对α_M求导：

$$\begin{aligned} \frac{\partial Loss}{\partial \alpha_M} &= \Sigma_{i=1}^n w_{Mi} [2I(y_i \neq g_M(x_i)) - 1] e^{\alpha_M [2I(y_i \neq g_M(x_i)) - 1]} \\ &= \Sigma_{y_i \neq g_M(x_i)} w_{Mi} e^{\alpha_M} - \Sigma_{y_i = g_M(x_i)} w_{Mi} e^{-\alpha_M} \\ &= 0 \end{aligned}$$

因此

$$\Sigma_{y_i \neq g_M(x_i)} w_{Mi} e^{\alpha_M} = \Sigma_{y_i = g_M(x_i)} w_{Mi} e^{-\alpha_M}$$

$$\begin{aligned} e^{2\alpha_M} &= \frac{\Sigma_{y_i = g_M(x_i)} w_{Mi}}{\Sigma_{y_i \neq g_M(x_i)}} \\ &= \frac{\Sigma_{i=1}^n w_{Mi} - \Sigma_{y_i \neq g_M(x_i)} w_{Mi}}{\Sigma_{y_i \neq g_M(x_i)}} \\ &= \frac{\Sigma_{i=1}^n w_{Mi} - \Sigma_{i=1}^n w_{Mi} I(y_i \neq g_M(x_i))}{\Sigma_{i=1}^n w_{Mi} I(y_i \neq g_M(x_i))} \\ &= \frac{1 - \frac{\Sigma_{i=1}^n w_{Mi} I(y_i \neq g_M(x_i))}{\Sigma_{i=1}^n w_{Mi}}}{\frac{\Sigma_{i=1}^n w_{Mi} I(y_i \neq g_M(x_i))}{\Sigma_{i=1}^n w_{Mi}}} \end{aligned}$$

而$\frac{\Sigma_{i=1}^n w_{Mi} I(y_i \neq g_M(x_i))}{\Sigma_{i=1}^n w_{Mi}}$ 正好是第M个弱学习器的加权训练误差err_M，因此 $\alpha_M = \frac{1}{2} log(\frac{1-err_M}{err_M})$. 由此我们也可计算出样本权重的更新公式

$$\begin{aligned} w_{(M+1)i} &= e^{-y_iF_{M-1}(x_i) - y_i \alpha_M g_M(x_i)} \\ &= w_{Mi} e^{-y_i \alpha_M g_M(x_i)} \end{aligned}$$

为了满足权重的定义，我们在每次得到新的样本权重后需要进行归一化处理。