Adaboost 训练误差界

- $\frac{1}{N}\sum I(G(x_i)\neq y_i) \le \frac{1}{N}\sum \exp\{-y_i f(x_i)\} = \prod_m Z_m$

证明：

当 $G(x_i)\neq y_i$ 时 $\exp\{-y_i f(x_i)\} \ge 1$ 前部分成立

$$D_{m+1} = (w_{m+1,1},\cdots,w_{m+1,i},\cdots,w_{m+1,N}) \tag{8.3}$$

$$w_{m+1,i} = \frac{w_{mi}}{Z_m}\exp(-\alpha_m y_i G_m(x_i)),\quad i=1,2,\cdots,N \tag{8.4}$$

这里，Z_m 是规范化因子

$$Z_m = \sum_{i=1}^{N} w_{mi}\exp(-\alpha_m y_i G_m(x_i)) \tag{8.5}$$

它使 D_{m+1} 成为一个概率分布.

$\Downarrow$

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

现推导如下：

$$\frac{1}{N}\sum \exp(-y_i f(x_i))$$
$$= \frac{1}{N}\sum_i \exp\{-\sum_m \alpha_m y_i G_m(x_i)\}$$
$$= \frac{1}{N}\sum_i \prod_m \exp\{-\alpha_m y_i G_m(x_i)\} \quad \rightarrow e^{x_1+x_2} = e^{x_1}e^{x_2}$$
$$= \sum_i w_{1i} \prod_m \exp\{-\alpha_m y_i G_m(x_i)\} \quad \Rightarrow \text{初始权重 } w_{1i} = \frac{1}{N}$$
$$= Z_1 \sum w_{2i} \prod_{m=2} \exp\{-\alpha_m y_i G_m(x_i)\}$$
$$= Z_1 \cdot Z_2 \cdot \sum w_{3i} \prod_{m=3} \exp\{-\alpha_m y_i G_m(x_i)\}$$
$$= \prod_m Z_m$$

$Z_1 = \dfrac{w_{1i}\exp(-\alpha_1 y_i G_1(x_i))}{w_{2,i}}$

$Z_2 w_{3,i} = w_{2i}\exp\{-\alpha_2 y_i G_2(x_i)\}$

- 这说明，可以在每一轮取适当 G_m 使 Z_m 最小，从而使训练误差下降最快。

对2分类问题

- $\prod_m Z_m = \prod_m [2\sqrt{e_m(1-e_m)}] = \prod_m \sqrt{(1-4\gamma_m^2)} \le \exp\{-2\sum_m \gamma_m^2\}$

  这里 $\gamma_m = \frac{1}{2}(1-e_m)$

△ 证明:

$Z_m = \sum_i w_{mi} \exp\{-\alpha_m y_i G_m(x_i)\}$

$\because$ 当 $y_i = G_m(x_i)$ 时

$y_i \cdot G_m(x_i) = 1$

当 $y_i \ne G_m(x_i)$ 时

$y_i \cdot G_m(x_i) = -1$

$\therefore Z_m = \sum_{y_i = G_m(x_i)} w_{mi} \exp\{-\alpha_m) + \sum_{y_i \ne G_m(x_i)} w_{mi} \exp(\alpha_m)$

注意到: $e_m = \sum w_{mi} I(G_m(x_i) \ne y_i)$.

$Z_m = (1-e_m)\exp(-\alpha_m) + e_m \exp\{\alpha_m)$

$\because \alpha_m = \frac{1}{2}\log\frac{1-e_m}{e_m}$

$\therefore \exp(-\alpha_m) = \exp\{-\frac{1}{2}\log\frac{1-e_m}{e_m}\}$

$= \exp\{\log\sqrt{\frac{e_m}{1-e_m}}\}$

$= \sqrt{\frac{e_m}{1-e_m}}$

$\exp\{\alpha_m\} = \sqrt{\frac{1-e_m}{e_m}}$

$Z_m = (1-e_m)\sqrt{\frac{e_m}{1-e_m}} + e_m\sqrt{\frac{1-e_m}{e_m}}$

$= 2\sqrt{e_m(1-e_m)}$

令 $\gamma_m = \frac{1}{2} - e_m$ $\quad\therefore e_m = \frac{1}{2} - \gamma_m$

$\therefore Z_m = 2\sqrt{(\frac{1}{2}-\gamma_m)(\frac{1}{2}+\gamma_m)}$

$= 2\sqrt{\frac{1}{4} - \gamma_m^2}$

$= \sqrt{1 - 4\gamma_m^2}$

△. $\prod \sqrt{1-4\gamma_m^2} \le \exp\{-2\sum \gamma_m^2\} = \prod \exp\{-2\gamma_m^2\}$

∴即证: $\sqrt{1-4\gamma_m^2} \le \exp\{-2\gamma_m^2\}$

即证: $1-4\gamma_m^2 \le \exp\{-4\gamma_m^2\}$

令 $x = 4\gamma_m^2$

∵ $\gamma_m = 1-e_m$ $\quad e_m \in [0,1]$

∴ $\gamma_m \in [-\frac{1}{2}, \frac{1}{2}]$ $\quad$ ∴ $x \in [0,1]$

构造 $f(x) = \exp\{-x) + x - 1$

$f'(x) = -\exp\{-x\} + 1 \ge 0$, $\quad x \in [0,1]$

∴ $f(x)$ ↑ 又∵ $f(0)=0$

∴ $f(x) \ge 0$ 在 $[0,1]$ 上成立

$\exp\{-x\} \ge 1-x$

即 $\exp\{-4\gamma_m^2\} \ge 1-4\gamma_m^2$

推论 • 如果存在 $\gamma > 0$, 对所有 m 有 $\gamma_m \ge \gamma$, 则

$\frac{1}{N}\sum I(G(x_i) \ne y_i) \le \exp\{-2M\gamma^2\}$

这表明, 在此条件下, Adaboost 的训练误差是指数速率下降,

因为 $\gamma_m = 1-e_m$, 误差越小, m 越大, 则误差上界以指数级减少.