II. Prove that AdaBoost.M1 is equal to forward stagewised additive modeling using the loss function. $L(y, f(x)) = \exp(-yf(x))$

Basis classifier : $G_m(x) \in \{-1, 1\}$

For the $G_m(x)$ and β_m to be added, one must solve

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

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

depend neither on β nor G

$$\Rightarrow (\beta_m, G_m(x)) = \arg\min_{\beta, G} \sum_{i=1}^{N} \boxed{w_i^{(m)}} \exp[-y_i \beta G(x_i)] \qquad (1.1)$$

$$= \arg\min_{\beta, G} e^{-\beta} \sum_{y_i = G(x_i)} w_i^{(m)} + e^{\beta} \sum_{y_i \neq G(x_i)} w_i^{(m)}$$

$I(y_i \neq G(x_i))$

$$= \arg\min_{\beta, G} e^{\beta} \sum_{i=1}^{N} w_i^{(m)} \cdot I(y_i \neq G(x_i)) + e^{-\beta} \sum_{i=1}^{N} w_i^{(m)} - e^{-\beta} \sum_{i=1}^{N} w_i^{(m)} \cdot$$

$$= \arg\min_{\beta, G} (e^{\beta} - e^{-\beta}) \sum_{i=1}^{N} w_i^{(m)} \cdot I(y_i \neq G(x_i)) + e^{-\beta} \sum_{i=1}^{N} w_i^{(m)}$$

$$= \arg\min_{\beta, G} \sum_{i=1}^{N} w_i^{(m)} \left[(e^{\beta} - e^{-\beta}) \frac{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i))}{\sum_{i=1}^{N} w_i^{(m)}} + e^{-\beta} \right]$$

use $error_m = \dfrac{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i^{(m)}}$ (mentioned in sec 1)

assum $f(\beta) = \sum_{i=1}^{N} w_i^{(m)} \cdot [(e^{\beta} - e^{-\beta}) error_m + e^{-\beta}]$

$f'(\beta) = \sum_{i=1}^{N} w_i^{(m)} \cdot [(e^{\beta} + e^{-\beta}) error_m + e^{-\beta}]$

$$\Rightarrow \sum_{i=1}^{N} w_i^{(m)} \left[(e^{\beta_m} + e^{-\beta_m}) \cdot error_m - e^{-\beta_m} \right] = 0$$

$$\Rightarrow (e^{\beta_m} + e^{-\beta_m}) \cdot error_m - e^{-\beta_m} = 0$$

$$\Rightarrow e^{2\beta_m} \cdot error_m + error_m - 1 = 0$$

$$\Rightarrow \beta_m = \frac{1}{2} \log \left(\frac{1 - error_m}{error_m} \right)$$

For each $\beta > 0$, $G_m = \arg\min_{G} \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i))$

the classifier that minimizes the weighted error rate

The approximation is then updated

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

AdaBoost.M1 minimizes the exponential loss criterion via a forward-stagewise additive modeling approach

The weights for the next iteration to be

$$w_i^{(m+1)} = w_i^{(m)} \cdot \exp[-\beta_m y_i G_m(x_i)]$$

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

$$= w_i^{(m)} \cdot \exp[2\beta_m I(y_i \neq G_m(x_i))] \cdot \exp(-\beta_m)$$

see Algorithm 1.

$$= \underbrace{w_i^{(m)} \exp[\alpha_m I(y_i \neq G_m(x_i))]}_{\text{is equal to Algorithm 1. 2d}} \cdot \underbrace{\exp(-\beta_m)}_{\text{no effect}} \quad \text{where } \boxed{\alpha_m} = 2\beta_m$$