Two-class classification problem

$y_i \in \{-1; 1\}$

x_i is the vector of the inputs

Ingredients $\sum_{i=1}^{N} L(y_i; f(x_i))$

loss function : $L(y; f(x)) = \exp\{-y f(x)\}$

forward stagewise additive modelling

$$(\beta_m, \gamma_m) = \arg\min_{\beta,\gamma} \sum_{i=1}^{N} L\Big(y_i, \underbrace{\hat{f}^{[m-1]}(x_i)}_{\text{current classifier}} + \overbrace{\beta}^{\text{weight}} \underbrace{b(x_i;\gamma)}_{\text{new weak classifier}}\Big)$$

AdaBoost: $\hat{f}^{[m-1]}(x_i) = \sum_{k=1}^{m-1} \hat{\beta}^{[k]} \hat{g}^{[k]}(x_i)$

at the m-th step

$$(\beta_m, g_m) = \arg\min_{\beta,g} \sum_{i=1}^{N} \exp\Big\{-y_i \sum_{k=1}^{m} \beta^{[k]} g(x_i)\Big\}$$

$$= \arg\min_{\beta,g} \sum_{i=1}^{N} \exp\Big\{-y_i \Big(\sum_{k=1}^{m-1} \hat{\beta}^{[k]} \hat{g}^{[k]}(x_i) + \beta g(x_i)\Big)\Big\}$$

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

where $w_i^{[m-1]} = \exp\{-y_i \hat{f}^{[m-1]}(x_i)\}$ does not depend neither on β nor on g (given by the previous iteration)

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

Two-step procedure
- first we minimize in g
- then " " " β

* if $y_i = g(x_i) \to y_i g(x_i) = 1$
if $y_i \neq g(x_i) \to y_i g(x_i) = -1$

$$g_m = \arg\min_{g} \sum_{i=1}^{N} w_i^{[m-1]} \exp\{-y_i \beta g(x_i)\}$$

$$\overset{*}{=} \arg\min_{g} \Big\{ \sum_{i: y_i = g(x_i)} w_i^{[m-1]} e^{-\beta} + \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{\beta} \Big\}$$

$$= \arg\min_{g} \Big\{ \sum_{i: y_i = g(x_i)} w_i^{[m-1]} e^{-\beta} + \underbrace{\sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{-\beta}}_{} - \underbrace{\sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{-\beta}}_{} + \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{\beta} \Big\}$$

$$= \arg\min_{g} \Big\{ \sum_{i=1}^{N} w_i^{[m-1]} e^{-\beta} + (e^{\beta} - e^{-\beta}) \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} \Big\}$$

$$= \arg\min_{g} \Big\{ \sum_{i=1}^{N} w_i^{[m-1]} e^{-\beta} + (e^{\beta} - e^{-\beta}) \sum_{i=1}^{N} w_i^{[m-1]} \mathbb{1}(y_i \neq g(x_i)) \Big\}$$

$$\Rightarrow g = \arg\min_{g} \Big\{ \underbrace{\sum_{i=1}^{N} w_i^{[m-1]} \mathbb{1}(y_i \neq g(x_i))}_{} \Big\}$$

minimizer of the weighted missclassification rate

$$\beta = \arg\min_{\beta} \sum_{i=1}^{N} w_i^{[m-1]} \exp\{-y_i \beta g(x_i)\}$$

$$y_i = g(x_i) \longrightarrow \sum_{i: y_i = g(x_i)} w_i^{[m-1]} e^{-\beta}$$

$$y_i \neq g(x_i) \longrightarrow \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{\beta}$$

$$L = \sum_{i: y_i = g(x_i)} w_i^{[m-1]} e^{-\beta} + \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{\beta}$$

$$\frac{\partial L}{\partial \beta} = \underbrace{-\sum_{i: y_i = g(x_i)} w_i^{[m-1]} e^{-\beta} + \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} e^{\beta} = 0}_{\text{multiply and divide by } e^{\beta}}$$

$$-\sum_{i: y_i = g(x_i)} w_i^{[m-1]} + e^{2\beta} \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} = 0$$

$$e^{2\beta} \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} = \sum_{i=1}^{N} w_i^{[m-1]} - \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]}$$

$$e^{2\beta} = \frac{\sum_{i=1}^{N} w_i^{[m-1]} \Big/ \sum_{i=1}^{N} w_i^{[m-1]} - \sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} \Big/ \sum_{i=1}^{N} w_i^{[m-1]}}{\sum_{i: y_i \neq g(x_i)} w_i^{[m-1]} \Big/ \sum_{i=1}^{N} w_i^{[m-1]}}$$

$$= \frac{1 - err}{err} \Rightarrow \beta = \frac{1}{2} \log \frac{1 - err}{err}$$

$$\text{where } err = \frac{\sum_{i=1}^{N} w_i^{[m-1]} \mathbb{1}(y_i \neq g(x_i))}{\sum_{i=1}^{N} w_i^{[m-1]}}$$

$$\alpha = 2\beta$$

$$\alpha = \log\left(\frac{1 - err}{err}\right)$$

$\hat{g}^{[m]}$ is the minimizer of the weighted missclassification rate

$$\hat{\beta}^{[m]} = \frac{1}{2} \log \frac{1 - err^{[m]}}{err^{[m]}}$$

$$\hat{\alpha}^{[m]} = \log \frac{1 - err^{[m]}}{err^{[m]}}$$

Our general classifier is updated as

$$\hat{f}^{[m]}(x) = \hat{f}^{[m-1]}(x) + \hat{\beta}^{[m]} \hat{g}^{[m]}(x)$$

so the weights for the next iteration

$$w_i^{[m]} = w_i^{[m-1]} \exp\left\{ \underline{-y_i \hat{\beta}^{[m]} \hat{g}^{[m]}(x_i)} \right\}$$

Since

$$-y_i \hat{g}^{[m]}(x_i) = -\sum_{i: y_i = g(x_i)} 1 - \sum_{i: y_i \neq g(x_i)} (-1)$$

$$= \underline{-\sum_{i: y_i = g(x_i)} 1} + \underline{\sum_{i: y_i \neq g(x_i)} 1 + \sum_{i: y_i \neq g(x_i)} 1} \underline{- \sum_{i: y_i \neq g(x_i)} 1}$$

$$= \underline{-1} + \underline{2 \sum_{i: y_i \neq g(x_i)} 1}$$

Plugging-in

$$w_i^{[m]} = w_i^{[m-1]} \exp\left\{ \hat{\beta}^{[m]} \left(2 \sum_{i: y_i \neq g(x_i)} 1 - 1 \right) \right\}$$

$$= w_i^{[m-1]} \exp\left\{ \underline{\hat{\beta}^{[m]} 2}\, \mathbb{1}(y_i \neq g(x_i)) - \hat{\beta}^{[m]} \right\}$$

$\alpha = 2\beta$

$$= \underline{w_i^{[m-1]}}\ \underbrace{e^{-\hat{\alpha}^{[m]}/2}}_{\text{constant}}\ \underline{e^{\hat{\alpha}^{[m]} \mathbb{1}(y_i \neq g(x_i))}}$$

we increase the weight of the missclassified observations