Boosting

1. Initialize weights $w_i = 1/N$
2. for $m = 1, 2, \ldots, M$

   a. fit a classifier $G^{(m)}(x)$

   $$G^{(m)}(x) = \underset{G}{\arg\inf} \sum_{i=1}^{n} w_i^{(m)} L(G(x_i), y_i)$$

   b. Compute $$Err_m = \frac{\sum_{i=1}^{n} w_i^{(m)} I(y_i \neq G^{(m)}(x_i))}{\sum_{i=1}^{n} w_i^{(m)}}$$

   c. Update $$\alpha_m = \ln\left(\frac{1 - err_m}{err_m}\right)$$

   d. Update $$w_i = w_i \exp\left\{\alpha_m I(y_i \neq G^{(m)}(x_i))\right\}$$

After convergence of the weights

3. Output $$G(x) = \text{Sign}\left\{\sum_{m=1}^{M} \alpha_m G^{(m)}(x)\right\}$$

Stagewise additive Models

1. Initialize $f_0(x) = 0$
2. For $m = 1, 2, \ldots, M$

   a. Compute

   $$(\beta_m, \gamma_m) = \underset{\beta, \gamma}{\arg\inf} \sum_{i=1}^{N} L(y_i, f^{(m-1)}(x_i) + \beta b(x_i; \gamma))$$

   b. Set $$f^{(m)}(x) = f^{(m-1)}(x) + \beta_m b(x_i; \gamma_m)$$

where $b(\cdot\,;\cdot)$ is a general set of basis functions.

Exponential Loss & Ada Boost

Claim: Ada Boost is equivalent to forward stagewise additive modeling

using $L(y, f(x)) = \exp(-y f(x))$

At each stage of f.s.a.m we solve

$$(\beta_m, G_m) = \underset{\beta, G}{\arg\min} \sum_{i=1}^{n} \exp\left\{-y_i f_{m-1}(x_i) + \beta G(x_i)\right\}$$

$$= \underset{\beta, G}{\arg\min} \sum_{i=1}^{n} w_i^{(m)} \exp\left(-y_i \beta G(x_i)\right)$$

$$= \underset{\beta, G}{\arg\inf} \left\{ e^{-\beta} \sum_{i=1}^{n} w_i^{(m)} I(y_i = G(x_i)) + e^{\beta} \sum_{i=1}^{n} w_i^{(m)} I(y_i \neq G(x_i)) \right\}$$

$$= \underset{\beta, G}{\arg\inf} \left\{ (e^{\beta} - e^{-\beta}) \underbrace{\sum_{i=1}^{n} w_i^{(m)} I(y_i \neq G(x_i))}_{A} + e^{-\beta} \underbrace{\sum_{i=1}^{n} w_i^{(m)}}_{C} \right\}$$

Minimization problem will be done in H.W.

Sol: $G_m = \underset{G}{\arg\min} \sum_{i=1}^{n} w_i^{(m)} I(y_i \neq G(x_i))$

$$\beta_m = \frac{1}{2} \log \frac{1 - err_m}{err_m}$$

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

$$w_i^{(m+1)} = w_i^{(m)} e^{-y_i \beta_m G_m(x)}$$

which is identical to the Ada-Boost weights.

Ex: Regression Trees

Define $f_m(x) = \sum_{m=1}^{m} T(x; \Theta_m)$ $\quad \Theta_m = \{R_j^{(m)}, \gamma_j^{(m)}\}_{j=1}^{J}$

$$T(x;\Theta) = \sum_{j=1}^{J} \gamma_j \, I(x \in R_j)$$

$$\hat{\gamma}_j^{(m)} = \underset{\gamma_j^{(m)}}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, f_{m-1}(x_i) + \gamma_j^{(m)})$$

$$\hat{\Theta}_m = \underset{\Theta_m}{\operatorname{argmin}} \sum_{i=1}^{N} w_i^{(m)} \exp\{-y_i T(x_i; \Theta_m)\}$$