L2 Boosting

Uses squared error loss

$$L(y_i, f_{m-1}(x_i) + \beta\phi(x_i;\gamma)) = (y_i - (f_{m-1}(x_i) + \beta\phi(x_i;\gamma)))^2$$

$$= (\underbrace{y_i - f_{m-1}(x_i)}_{\substack{r_{im} \\ \text{(current residual)}}} - \beta\phi(x_i;\gamma))^2$$

assume $\beta = 1$

$$= (r_{im} - \phi(x_i;\gamma))^2$$

AdaBoost

Uses exponential loss function

$$L_m(\phi) = \sum_{i=1}^{N} \exp\left[-\tilde{y}_i f_m(x_i)\right]$$

$$= \sum_{i=1}^{N} \exp\left[-\tilde{y}_i \left(f_{m-1}(x_i) + \beta\, \phi(x_i)\right)\right]$$

$$= \sum_{i=1}^{N} \exp\left[-\tilde{y}_i \left(f_{m-1}(x_i)\right) - \tilde{y}_i\, \beta\, \phi(x_i)\right]$$

$$= \sum_{i=1}^{N} \left[\underbrace{\exp\left[-\tilde{y}_i \left(f_{m-1}(x_i)\right)\right]}_{w_{i,m}} \cdot \exp\left[-\beta \tilde{y}_i\, \phi(x_i)\right]\right]$$

is a weight applied to data instance i.

$$= \sum_{i=1}^{N} w_{i,m} \cdot \exp\left[-\beta \cdot \tilde{y}_i \phi(x_i)\right]$$

$\tilde{y}_i \cdot \phi(x_i) = 1$ if the actual and the predicted outputs are the same, otherwise it is -1.

$$= e^{-\beta} \sum_{\tilde{y}_i = \phi(x_i)} w_{i,m} + e^{\beta} \sum_{\tilde{y}_i \neq \phi(x_i)} w_{i,m}$$

$$= e^{-\beta} \left(\sum_{i=1}^{N} w_{i,m} - \sum_{\tilde{y}_i \neq \phi(x_i)} w_{i,m} \right) + e^{\beta} \sum_{\tilde{y}_i \neq \phi(x_i)} w_{i,m}$$

$$= \left(e^{\beta} - e^{-\beta}\right) \sum_{\tilde{y}_i \neq \phi(x_i)} w_{i,m} + e^{-\beta} \sum_{i=1}^{N} w_{i,m}$$

$$= \left(e^{\beta} - e^{-\beta}\right) \sum_{i=1}^{N} w_{i,m} \, \mathbb{I}\left(\tilde{y}_i \neq \phi(x_i)\right) + e^{-\beta} \sum_{i=1}^{N} w_{i,m}$$