Chapter 16
Adaptive basis function models

16.1 AdaBoost

16.1.1 Representation

$$y = \mathrm{sign}(f(\boldsymbol{x})) = \mathrm{sign}\left(\sum_{i=1}^{m} \alpha_m G_m(\boldsymbol{x})\right) \tag{16.1}$$

where $G_m(\boldsymbol{x})$ are sub classifiers.

16.1.2 Evaluation

$L(y, f(\boldsymbol{x})) = \exp[-yf(\boldsymbol{x})]$ i.e., exponential loss function

$$(\alpha_m, G_m(x)) = \arg\min_{\alpha,G} \sum_{i=1}^{N} \exp\left[-y_i(f_{m-1}(\boldsymbol{x}_i) + \alpha G(\boldsymbol{x}_i))\right] \tag{16.2}$$

Define $\bar{w}_{mi} = \exp\left[-y_i(f_{m-1}(\boldsymbol{x}_i)\right]$, which is constant w.r.t. α, G

$$(\alpha_m, G_m(x)) = \arg\min_{\alpha,G} \sum_{i=1}^{N} \bar{w}_{mi} \exp\left(-y_i \alpha G(x_i)\right) \tag{16.3}$$

16.1.3 Optimization

16.1.3.1 Input

$$\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y2), \ldots, (\boldsymbol{x}_N, y_N)\}$$
$$\text{where } \boldsymbol{x}_i \in \mathbb{R}^D,\ y_i \in \{-1, +1\}$$
$$\text{Weak classifiers } \{G_1, G_2, \ldots, G_m\}$$

16.1.3.2 Output

Final classifier: $G(x)$

16.1.3.3 Algorithm

1. Initialize the weights' distribution of training data(when $m = 1$)

$$\mathcal{D}_0 = (w_{11}, w_{12}, \cdots, w_{1n}) = (\frac{1}{N}, \frac{1}{N}, \cdots, \frac{1}{N})$$

2. Iterate over $m = 1, 2, \ldots, M$
   (a) Use training data with current weights' distribution $\mathcal{D}_m$ to get a classifier $G_m(\boldsymbol{x})$
   (b) Compute the error rate of $G_m(\boldsymbol{x})$ over the training data

$$e_m = P(G_m(\boldsymbol{x}_i) \neq y_i) = \sum_{i=1}^{N} w_{mi} \mathbb{I}(G_m(\boldsymbol{x}_i) \neq y_i) \tag{16.4}$$

   (c) Compute the coefficient of classifier $G_m(x)$

$$\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \tag{16.5}$$

   (d) Update the weights' distribution of training data

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(\boldsymbol{x}_i)) \tag{16.6}$$

   where Z_m is the normalizing constant

$$Z_m = \sum_{i=1}^{N} w_{mi} \exp(-\alpha_m y_i G_m(\boldsymbol{x}_i)) \tag{16.7}$$

3. Ensemble M weak classifiers

$$G(x) = \mathrm{sign} f(\boldsymbol{x}) = \mathrm{sign}\left[\sum_{m=1}^{M} \alpha_m G_m(\boldsymbol{x})\right] \tag{16.8}$$

16.1.4 The upper bound of the training error of AdaBoost

Theorem 16.1. *The upper bound of the training error of AdaBoost is*

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(G(\boldsymbol{x}_i) \neq y_i) \leq \frac{1}{N} \sum_{i=1}^{N} \exp(-y_i f(\boldsymbol{x}_i)) = \prod_{m=1}^{M} Z_m \tag{16.9}$$

Note: the following equation would help proof this theorem

$$w_{mi} \exp(-\alpha_m y_i G_m(\boldsymbol{x}_i)) = Z_m w_{m+1,i} \tag{16.10}$$