Boosted Learning

Boosting is an iterative algorithm that iteratively selects and adds the best decision functions to a weighted committee.

Assume training data composed of M sample observations $\{\vec{X}_m\}$ where each sample observation is labeled with an indicator variable $\{y_m\}$

$y_m = +1$ for examples of the target pattern (class 1)
$y_m = -1$ for all other examples (class 2)

The algorithm estimates a weight for each training sample, $b_m^{(i)}$. The weights are initially set to 1.

The Algorithm

1) Initialisation.

$$i \leftarrow 0;\; n \leftarrow 1; \quad \forall_{m=1}^{M} b_m = 1; \qquad \alpha_1 \leftarrow 1;$$

$$\forall h_s() \in \{h_s()\} : E_s\left(h_s(\vec{X}_m)\right) = \sum_{m=1}^{M} I\left(h_s(\vec{X}_m)\right)$$

$$h_1(\vec{X}) = \arg\!-\!\min_{h_s(\vec{X})} \left\{ \sum_{m=1}^{M} I\left(h_s(\vec{X}_m)\right) \right\}$$

where $I\left(h_s(\vec{X}_m)\right)$ is an error function that allows us to count the number of errors made by a decision function with the training data.

$$I\left(h_s(\vec{X}_m)\right) = \begin{cases} 1 & h_s(\vec{X}_m) < 0 \\ 0 & h_s(\vec{X}_m) \geq 0 \end{cases}$$

2) Loop until E_i < Specified maximum Error rate.

$$i \leftarrow i+1$$

$$h_i(\vec{X}) = \arg\!-\!\min_{h_s(\vec{X})} \left\{ \sum_{m=1}^{M} b_m \cdot I\left(y_m \cdot \left(h_s(\vec{X}_m) + \sum_{n=1}^{i-1} \alpha_n h_n(\vec{X}_m) \right) \right) \right\}$$

$$E_i = \frac{\sum_{m=1}^{M} b_m \cdot I\left(y_m \cdot \left(h_i(\vec{X}_m) + \sum_{n=1}^{i-1} \alpha_n h_n(\vec{X}_m)\right)\right)}{\sum_{m=1}^{M} b_m}$$

$$\alpha_i = \log\left(\frac{1-E_i}{E_i}\right)$$

$$\forall_{m=1}^{M}: \; b_m^{(i-1)} \cdot e^{\alpha_i I(y_m \cdot F_i(\vec{X}_m))}$$

end Loop. set $N \leftarrow i$; $\{\alpha_i\}$, $\{h_i(-)\}$

Explanation for each step

Initialisation.

a) initialize the index, i, the weights for training data b_m and for the first classifier α_1

$$n \leftarrow 1; \qquad \forall_{m=1}^{M} b_m = 1; \qquad \alpha_1 \leftarrow 1;$$

b) Choose the decision function, $h_s()$, that gives the lowest error rate with the training data .

$$h_1(\vec{X}) = \arg\!-\!\min_{h_s(\vec{X})}\left\{\sum_{m=1}^{M} I\left(h_s(\vec{X}_m)\right)\right\}$$

where $I\left(h_s(\vec{X}_m)\right)$ is an error function that allows us to count the number of errors made by each decision function

$$I\left(h_s(\vec{X}_m)\right) = \begin{cases} 1 & h_s(\vec{X}_m) < 0 \\ 0 & h_s(\vec{X}_m) \geq 0 \end{cases}$$

2) Loop until E_i < Specified maximum Error rate.

a) Update the index variable $i \leftarrow i+1$