The AdaBoost algorithm

Input to AdaBoost: m labelled examples $S = (x_1, y_1), \ldots, (x_m, y_m)$ where each label $y_i \in \pm 1$

Notation:

- $\mathcal{D}_t$ denotes the t-th distribution Adaboost constructs over the m examples. $\mathcal{D}_t(i)$ denotes $\Pr_{\mathcal{D}_t}(x_i)$.
- h_t is the t-th hypothesis. ± 1 valued
- ϵ_t denotes $\Pr_{i \in \mathcal{D}_t}[h_t(x_i) \neq y_i]$ the error of h_t w.r.t. $\mathcal{D}_t$ (alg. computes this)

The algorithm:

1. Initialize $\mathcal{D}_1(i) = \frac{1}{m}$ for each $i = 1, \ldots, m$. (unif. over S: first dist. used.)
2. For $t = 1$ to T do:
   (a) Run weak learner L on $\mathcal{D}_t$ to get hypothesis h_t which has error ϵ_t w.r.t. $\mathcal{D}_t$.
   (b) Let $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$.
   (c) Update (new dist!)

$$\mathcal{D}_{t+1}(i) = \frac{\mathcal{D}_t(i) \cdot \exp(-\alpha_t \overbrace{y_i h_t(x_i)}^{\begin{cases} = +1 & \text{if } h_t(x_i) = y_i \\ -1 & \text{if } h_t(x_i) \neq y_i \end{cases}})}{Z_t}$$

   where Z_t is a normalization factor so that $\sum_{i=1}^m \mathcal{D}_{t+1}(i) = 1$.

   $Z_t = \sum_{i=1}^{m}$ (numerator)

3. Final hypothesis is $H(x) = \text{sign}(f(x))$ where $f(x) = \sum_{t=1}^T \alpha_t h_t(x)$.

α_t: weighting factor for h_t

$\text{sign}(f(x))$ → weighted majority vote

Ex: $\varepsilon_t = .4 \Rightarrow \alpha_t = .202$

$\varepsilon_t = .01 \Rightarrow \alpha_t = 2.29$