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. → what weak learner gives when we run it on $\mathcal{D}_t$
- ϵ_t denotes $\Pr_{i \in \mathcal{D}_t}[h_t(x_i) \neq y_i]$ the error of h_t w.r.t. $\mathcal{D}_t$ $= \sum_{i \in [m]: h_t(x_i) \neq y_i} \mathcal{D}_t(i)$

The algorithm:

1. Initialize $\mathcal{D}_1(i) = \frac{1}{m}$ for each $i = 1, \ldots, m$.
2. For $t = 1$ to T do:

   you compute this

   (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

   $$\mathcal{D}_{t+1}(i) = \frac{\mathcal{D}_t(i) \cdot \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

   where Z_t is a normalization factor so that $\sum_{i=1}^{m} \mathcal{D}_{t+1}(i) = 1$.
3. Final hypothesis is $H(x) = \text{sign}(f(x))$ where $f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$.

weighted sum

weighted maj vote of weak hyp's.