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:**

- $D_t$ denotes the $t$-th distribution AdaBoost constructs over the $m$ examples. $D_t(i)$ denotes $\Pr_{D_t}(x_i)$.
- $h_t$ is the $t$-th hypothesis.
- $\epsilon_t$ denotes $\Pr_{i \in D_t}[h_t(x_i) \neq y_i]$ the error of $h_t$ w.r.t. $D_t$

**The algorithm:**

1. Initialize $D_1(i) = \frac{1}{m}$ for each $i = 1, \ldots, m$.

2. For $t = 1$ to $T$ do:
   
   (a) Run weak learner $L$ on $D_t$ to get hypothesis $h_t$ which has error $\epsilon_t$ w.r.t. $D_t$.

   (b) Let $\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right)$

   (c) Update
   $$D_{t+1}(i) = \frac{D_t(i) \cdot \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$
   where $Z_t$ is a normalization factor so that $\sum_{i=1}^{m} D_{t+1}(i) = 1$.

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