Ada boost

$h \in \{\pm 1\}$, $y_i \in \{\pm 1\}$

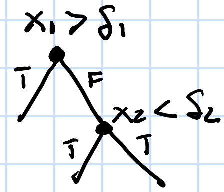
Input: $S = \{(x_i, y_i)\}_{i=1}^{n}$, hypothesis class $\mathcal{H}$, rounds T

Initialize: $D_1(i) = 1/n$ $\quad \sum_{i=1}^{n} D_t(i) = 1$

* h_t is a "rule of thumb"

for $t = 1 \ldots T$ **do**

① $h_t = \text{WeakLearn}(S, D_t, \mathcal{H})$ $\qquad h_t(x) \in \{\pm 1\}$



② $\varepsilon_t = \sum_{i:\,\text{wrong}} D_t(i)$

③ Choose $\alpha_t = \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t}$ $\qquad$ Adaboost.M2

④ $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \exp(-\alpha_t h_t(x_i) y_i)$

Output: $H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

Goal: Our objective is to understand how this "learning" procedure can achieve "small" error rates, but we will need to show some intermediate results

$\Longrightarrow$ New Material

Homework #3 Problem #2 Solution

Primal

$$\underset{w_T,\, \xi}{\arg\min} \left\{ \frac{1}{2} \|w_T\|_2^2 + C \sum_{i=1}^{n} \xi_i - B\, w_T^T w_s \right\}$$

$$\text{s.t.} \quad y_i (w_T^T x_i + b) \geq 1 - \xi_i \qquad \forall i$$

$$\xi_i \geq 0 \qquad \forall i$$

Dual

$$\underset{\alpha_i}{\arg\max} \left\{ \sum_{i=1}^{n} \alpha_i (1 - \beta\, y_i w_s^T x_i) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \right\}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

Let us look @ Z_t

$$Z_t = \sum_{i=1}^{n} D_t(i)\, e^{-\alpha_t h_t(x_i) y_i}$$

$$= e^{-\alpha_t} \overbrace{\boxed{\sum_{i:\text{correct}} D_t(i)}}^{1-\varepsilon_t} + e^{\alpha_t} \overbrace{\boxed{\sum_{i:\text{wrong}} D_t(i)}}^{\varepsilon_t}$$

$$= (1-\varepsilon_t)\, e^{-\alpha_t} + \varepsilon_t\, e^{\alpha_t}$$

* side note

$$e^{-\alpha_t} = e^{-\frac{1}{2}\log\frac{1-\varepsilon_t}{\varepsilon_t}} = e^{\log\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}} = \boxed{\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}}$$

$$e^{\alpha_t} = \boxed{\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}}$$

$$Z_t = (1-\varepsilon_t)\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} + \varepsilon_t\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}$$

$$= 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$$

What is h_t's error on D_{t+1}? $\quad \sum_{i:h_t(x_i)\neq y_i} D_{t+1}(i)$

$$\sum_{i: h_t(x_i) \neq y_i} D_{t+1}(i) = \frac{e^{\alpha_t}}{Z_t} \sum_{i: \text{wrong}} D_t(i)$$

$$= \frac{\varepsilon_t e^{\alpha_t}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} = \frac{\varepsilon_t \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}$$

$$\boxed{= \frac{1}{2}}$$

$$Z_t = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$$

What is the error of Adaboost?

$$\widehat{err}(H) = \frac{1}{n} \sum_{i=1}^{n} [\![H(x_i) \neq y_i]\!] \leq \{ - \}$$

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \exp(-\alpha_t h_t(x_i) y_i)$$

$$= \frac{D_{t-1}(i)}{Z_t Z_{t-1}} \exp(-\alpha_t h_t(x_i) y_i) \exp(-\alpha_{t-1} h_{t-1}(x_i) y_i)$$

$$\cdots$$

$$= \frac{1}{n \prod_{r=1}^{t} Z_r} \exp\left(-\sum_{r=1}^{t} \alpha_r y_i h_r(x_i)\right)$$