CS 6316 Homework 5

James Harbour
gtr8rh@virginia.edu

April 7, 2024

1 Problem 1

We have informally argued that the AdaBoost algorithm uses the weighting mechanism to "force" the weak learner to focus on the problematic examples in the next iteration. In this question, we will find some rigorous justification for this argument.

Show that the error of h_t w.r.t. the distribution $D^{(t+1)}$ is exactly $1/2$. That is, show that for every $t \in [T]$,

$$\sum_{i=1}^{m} D_i^{(t+1)} \mathbb{1}_{[y_i \neq h_t(x_i)]} = 1/2.$$

Proof. The error of h_t w.r.t. the distribution $D^{(t)}$ is given by

$$\varepsilon_t = \sum_{i=1}^{m} D_i^{(t)} \mathbb{1}_{[y_i \neq h_t(x_i)]},$$

whence the weight at round t given in the AdaBoost algorithm to the hypothesis h_t is

$$w_t = \frac{1}{2} \log\left(\frac{1}{\varepsilon_t} - 1\right).$$

As an intermediate computation, we note that

$$e^{-w_t} = e^{-\frac{1}{2}\log\left(\frac{1}{\varepsilon_t}-1\right)} = \left(\frac{1}{\varepsilon_t} - 1\right)^{-\frac{1}{2}} = \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} \tag{1}$$

$$\begin{aligned}
\sum_{i=1}^{m} D_i^{(t+1)} \mathbb{1}_{[y_i \neq h_t(x_i)]} &= \sum_{i=1}^{m} \frac{D_i^{(t)} e^{-w_t y_i h_t(x_i)}}{\sum_{j=1}^{m} D_j^{(t)} e^{-w_t y_j h_t(x_j)}} \mathbb{1}_{[y_i \neq h_t(x_i)]} \\
&\overset{(1)}{=} \frac{1}{\sum_{j=1}^{m} D_j^{(t)} \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{\,y_j h_t(x_j)}} \sum_{i=1}^{m} D_i^{(t)} \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{\,y_i h_t(x_i)} \mathbb{1}_{[y_i \neq h_t(x_i)]}
\end{aligned} \tag{2}$$

We treat the numerator of this expression first. Observe that

$$\begin{aligned}
\sum_{i=1}^{m} D_i^{(t)} \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{\,y_i h_t(x_i)} \mathbb{1}_{[y_i \neq h_t(x_i)]} &= \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{\,-1} \sum_{i=1}^{m} D_i^{(t)} \mathbb{1}_{[y_i \neq h_t(x_i)]} \\
&= \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} \cdot \varepsilon_t = \sqrt{\varepsilon_t(1-\varepsilon_t)}
\end{aligned}$$

To evaluate the denominator of the aforementioned expression, we use the triviality that $1 = \mathbb{1}_{[y_i=h_t(x_i)]} + \mathbb{1}_{[y_i\neq h_t(x_i)]}$ to expand

$$\begin{aligned}
\sum_{j=1}^{m} D_j^{(t)} \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{y_j h_t(x_j)} &= \sum_{j=1}^{m} D_j^{(t)} \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{y_j h_t(x_j)} \mathbb{1}_{[y_i\neq h_t(x_i)]} + \sum_{j=1}^{m} D_j^{(t)} \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{y_j h_t(x_j)} \mathbb{1}_{[y_i=h_t(x_i)]} \\
&= \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}^{-1} \cdot \sum_{j=1}^{m} D_j^{(t)} \mathbb{1}_{[y_i\neq h_t(x_i)]} + \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} \cdot \sum_{j=1}^{m} D_j^{(t)} \mathbb{1}_{[y_i=h_t(x_i)]} \\
&= \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} \cdot \varepsilon_t + \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} \cdot \sum_{j=1}^{m} D_j^{(t)} \mathbb{1}_{[y_i=h_t(x_i)]} \\
&= \sqrt{\varepsilon_t(1-\varepsilon_t)} + \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} \cdot \left(\sum_{j=1}^{m} D_j^{(t)} - \sum_{j=1}^{m} D_j^{(t)} \mathbb{1}_{[y_i\neq h_t(x_i)]} \right) \\
&= \sqrt{\varepsilon_t(1-\varepsilon_t)} + \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} \cdot (1-\varepsilon_t) = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}
\end{aligned}$$

Substituting these computations into equation (2), we find

$$\sum_{i=1}^{m} D_i^{(t+1)} \mathbb{1}_{[y_i\neq h_t(x_i)]} = \frac{1}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} \cdot \sqrt{\varepsilon_t(1-\varepsilon_t)} = \frac{1}{2}$$

as desired. □

2 Problem 2

In this exercise we discuss the VC-dimension of classes of the form $L(B,T)$. We proved an upper bound of $O(dT\log(dT))$, where $d = \mathrm{VCdim}(B)$. Here we wish to prove an almost matching lower bound. However, that will not be the case for all classes B.

Note that for every class B and every number $T \geq 1$, $\mathrm{VCdim}(B) \leq \mathrm{VCdim}(L(B,T))$. Find a class B for which $\mathrm{VCdim}(B) = \mathrm{VCdim}(L(B,T))$ for every $T \geq 1$. (*Hint*: Take $\mathcal{X}$ to be a finite set.)

Proof. Take $\mathcal{X} = \{1,\ldots,n\}$ and $B = 2^{\mathcal{X}}$. Suppose $T \geq 1$. If $f : \mathcal{X} \to \{\pm 1\}$ as any function, then for $w = (1,0,\ldots,0) \in \mathbb{R}^T$ and $h_t = f$ for $1 \leq t \leq T$, we have that

$$f(x) = \mathrm{sign}(f(x)) = \mathrm{sign}\left(\sum_{t=1}^{T} w_t f(x)\right)$$

whence $f \in L(B,T)$. So $B \subseteq L(B,T) \subseteq 2^{\mathcal{X}} = B$, whence $B = L(B,T)$ and consequently they have equal VC-dimensions. □