MLT Homework 14

Ana Borovac
Bas Haver

January 3, 2019

Question 1

Solution

We are going to use the hint, so fist divide samples into k chunks $S_1, \ldots, S_k$, each of size $m_{\mathcal{H}}(\epsilon)$. Now, we would like to prove:

$$P(\forall i \in \{1, \ldots, k\} : L_D(A(S_i)) > \min_{h \in \mathcal{H}} L_D(h) + \epsilon) \leq \delta_0^k$$

We know that for each S_i it holds:

$$P(L_D(A(S_i)) > \min_{h \in \mathcal{H}} L_D(h) + \epsilon) \leq \delta_0$$

Therefore:

$$\begin{aligned}
&P(\forall i \in \{1, \ldots, k\} : L_D(A(S_i)) > \min_{h \in \mathcal{H}} L_D(h) + \epsilon) = \\
&= \prod_{i=1}^{k} P(L_D(A(S_i)) > \min_{h \in \mathcal{H}} L_D(h) + \epsilon) \\
&\leq \delta_0^k
\end{aligned}$$

Question 2

Prove that the function h given in Equation (10.5) equals the piece-wise constant function defined according to the same thresholds as h.

Solution

We would like to prove that is holds if $x \in (\theta_{k-1}, \theta_k]$, then $h(x) = (-1)^k = g(x)$.

$$
\begin{aligned}
h(x) &= \operatorname{sign}\left(\sum_{t=1}^{T} w_t \operatorname{sign}(x - \theta_{t-1})\right) \\
&= \operatorname{sign}(\frac{1}{2}\operatorname{sign}(x-\theta_0) + (-1)^2\operatorname{sign}(x-\theta_1)+ \\
&\quad + \cdots + \\
&\quad (-1)^k\operatorname{sign}(x-\theta_{k-1}) + (-1)^{k+1}\operatorname{sign}(x-\theta_k)+ \\
&\quad + \cdots + \\
&\quad + (-1)^T\operatorname{sign}(x-\theta_{T-1}))
\end{aligned}
$$

Since $x \in (\theta_{k-1}, \theta_k]$, it holds:

$$
\operatorname{sign}(x-\theta_j) = \begin{cases} 1; & j \in \{0, \ldots, k\} \\ -1; & j \in \{k+1, \ldots, T\} \end{cases}
$$

It follows:

$$
\begin{aligned}
h(x) = \operatorname{sign}(&\frac{1}{2} + (-1)^2 + \\
&+ \cdots + \\
&(-1)^k + (-1)^{k+1}(-1) + \\
&+ \cdots + \\
&+ (-1)^T(-1)) \\
= \operatorname{sign}&\left(\frac{1}{2} + (-1)^2 + \cdots + (-1)^k + (-1)^{k+2} + \cdots + (-1)^{T+1}\right) \\
= \operatorname{sign}&\left(\frac{1}{2} + \sum_{i=2}^{T+1}(-1)^i - (-1)^{k+1}\right) \\
= \operatorname{sign}&\left(\frac{1}{2} + \sum_{i=2}^{T+1}(-1)^i + (-1)^k\right) \\
= \operatorname{sign}&\begin{cases} \frac{1}{2} + 1 + (-1)^k; & T \text{ is odd} \\ \frac{1}{2} - 1 + (-1)^k; & T \text{ is even} \end{cases} \\
= \operatorname{sign}&\begin{cases} \frac{3}{2} + (-1)^k; & T \text{ is odd} \\ -\frac{1}{2} + (-1)^k; & T \text{ is even} \end{cases} \\
= \operatorname{sign}&\begin{cases} \frac{3}{2} + (-1); & T \text{ is odd}, k \text{ is odd} \\ \frac{3}{2} + 1; & T \text{ is odd}, k \text{ is even} \\ -\frac{1}{2} + (-1); & T \text{ is even}, k \text{ is odd} \\ -\frac{1}{2} + 1; & T \text{ is even}, k \text{ is even} \end{cases} \\
= \operatorname{sign}&\begin{cases} -\frac{1}{2}; & T \text{ is odd}, k \text{ is odd} \\ \frac{5}{2}; & T \text{ is odd}, k \text{ is even} \\ -\frac{3}{2}; & T \text{ is even}, k \text{ is odd} \\ \frac{1}{2}; & T \text{ is even}, k \text{ is even} \end{cases} \\
= &\begin{cases} -1; & k \text{ is odd} \\ 1; & k \text{ is even} \end{cases} \\
= &(-1)^k
\end{aligned}
$$

Question 3

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

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

Solution

From definition of $D_i^{(t+1)}$:

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

Now, we are going to use:

$$\begin{aligned}
y_i = h_t(x_i) \quad &\Rightarrow \quad y_i h_t(x_i) = 1 \\
y_i \neq h_t(x_i) \quad &\Rightarrow \quad y_i h_t(x_i) = -1
\end{aligned}$$

It follows:

$$\begin{aligned}
&= \frac{\sum_{i=1}^{m} D_i^{(t)} e^{-w_t y_i h_t(x_i)} \mathbb{1}_{[y_i \neq h_t(x_i)]}}{\sum_{j=1}^{m} D_j^{(t)} e^{-w_t y_j h_t(x_j)} \mathbb{1}_{[y_i \neq h_t(x_i)]} + \sum_{j=1}^{m} D_j^{(t)} e^{-w_t y_j h_t(x_j)} \mathbb{1}_{[y_i = h_t(x_i)]}} \\
&= \frac{e^{w_t} \sum_{i=1}^{m} D_i^{(t)} \mathbb{1}_{[y_i \neq h_t(x_i)]}}{e^{w_t} \sum_{j=1}^{m} D_j^{(t)} \mathbb{1}_{[y_i \neq h_t(x_i)]} + e^{-w_t} \sum_{j=1}^{m} D_j^{(t)} \mathbb{1}_{[y_i = h_t(x_i)]}} \\
&= \frac{e^{w_t} \epsilon_t}{e^{w_t} \epsilon_t + e^{-w_t}(1-\epsilon_t)}; \quad \epsilon_t = \sum_{i=1}^{m} D_i^{(t)} \mathbb{1}_{[y_i \neq h_t(x_i)]} \\
&= \frac{\epsilon_t}{\epsilon_t + e^{-2w_t}(1-\epsilon_t)} \\
&= \frac{\epsilon_t}{\epsilon_t + e^{-2\frac{1}{2}\log\left(\frac{1}{\epsilon_t}-1\right)}(1-\epsilon_t)}; \quad w_t = \frac{1}{2}\log\left(\frac{1}{\epsilon_t} - 1\right) \\
&= \frac{\epsilon_t}{\epsilon_t + \frac{\epsilon_t}{1-\epsilon_t}(1-\epsilon_t)} \\
&= \frac{\epsilon_t}{\epsilon_t + \epsilon_t} \\
&= \frac{1}{2}
\end{aligned}$$

Question 4

Solution