Homework No. 5

Alexander Shender 328626114

Snir Hordan 205689581

Question 1.

1. Prove that when running AdaBoost, the distribution is updated such that the error of the chosen weak classifier h_t, w.r.t the updated distribution $D_i^{(t+1)}$, is exactly $\frac{1}{2}$.

   That is, prove that $\sum_i D_i^{(t+1)} \cdot \mathbf{1}_{h_t(x_i)\neq y_i} = \frac{1}{2}$.

   Hint: You can fill the missing steps in the following derivation:

   $\sum_i D_i^{(t+1)} \cdot \mathbf{1}_{h_t(x_i)\neq y_i} = \cdots = \frac{\epsilon_t}{\epsilon_t+(1-\epsilon_t)\exp\{-2w_t\}} = \cdots = \frac{1}{2}$.

We start by indeed from writing the expression for the updated error value with the updated data weights (distribution) for the last chosen weak classifier:

$$\mathrm{E}_{t+1} = \sum_i D_i^{t+1} \cdot 1_{h_t(x_i)\neq y_i}$$

By using:

$$D_i^{t+1} = D_i^t \cdot \frac{\exp(-w_t y_i h_t(x_i))}{\sum_j D_j^t \exp\left(-w_t y_j h_t(x_j)\right)} = D_i^t \cdot \frac{\exp(-w_t y_i h_t(x_i))}{Z_t}$$

Putting back:

$$\mathrm{E}_{t+1} = \sum_i D_i^{t+1} \cdot 1_{h_t(x_i)\neq y_i} = \sum_i \frac{D_i^t \exp(-w_t y_i h_t(x_i))}{Z_t} \cdot 1_{h_t(x_i)\neq y_i}$$

Z_t is a normalization factor, so we can put it outside of the sum:

$$\mathrm{E}_{t+1} = \frac{\sum_i D_i^t \exp(-w_t y_i h_t(x_i))}{Z_t} \cdot 1_{h_t(x_i)\neq y_i}$$

We divide into 2 cases:

$$\begin{cases} E : h_t(x_i) = y_i \\ C : h_t(x_i) \neq y_i \end{cases}$$

For each case,

$$\begin{cases} E : 1_{h_t(x_i)\neq y_i} = 0\,;\ y_i h_t(x_i) = 1 \\ C : 1_{h_t(x_i)\neq y_i} = 1\,;\ y_i h_t(x_i) = -1 \end{cases}$$

So we get:

$$\mathrm{E}_{t+1} = \frac{\sum_{i\in E} D_i^t \exp(-w_t y_i h_t(x_i))}{Z_t} \cdot 0 + \frac{\sum_{i\in C} D_i^t \exp(-w_t y_i h_t(x_i))}{Z_t} \cdot 1 = \frac{\sum_i D_i^t \exp(w_t)}{Z_t}$$

The numerator contains the expression for the error value, since we have isolated for case $\{E:\ h_t(x_i) = y_i\}$:

$$\mathrm{E}_t = \sum_i D_i^t$$

$$\mathrm{E}_{t+1} = \frac{\mathrm{E}_t \exp(w_t)}{Z_t}$$

Opening the denominator using same 2 cases:

$$Z_t = \sum_{i \in E} D_i^t \exp(-w_t) + \sum_{i \in C} D_i^t \exp(w_t)$$

Putting back:

$$\mathrm{E}_{t+1} = \frac{\mathrm{E}_t \exp(w_t)}{\sum_{i \in E} D_i^t \exp(-w_t) + \sum_{i \in C} D_i^t \exp(w_t)} \quad ; \ / \exp(w_t)$$

$$\mathrm{E}_{t+1} = \frac{\mathrm{E}_t}{\sum_{i \in E} D_i^t \exp(-2w_t) + \sum_{i \in C} D_i^t} = \frac{\mathrm{E}_t}{\sum_{i \in E} D_i^t \exp(-2w_t) + \mathrm{E}_t}$$

Taking $\exp(-2w_t)$ out of the sum:

$$\mathrm{E}_{t+1} = \frac{\mathrm{E}_t}{\exp(-2w_t) \sum_{i \in E} D_i^t + \mathrm{E}_t}$$

The sum over weights of the correct predictions E, is 1-sum over incorrect predictions, since they sum to 1:

$$\sum_{i \in E} D_i^t = 1 - \sum_{i \in C} D_i^t = 1 - E_t$$

So we get:

$$\boxed{\mathrm{E}_{t+1} = \frac{\mathrm{E}_t}{\exp(-2w_t)\,(1 - E_t) + \mathrm{E}_t}}$$

Using the expression for the weight of the weak classifier:

$$w_t = \frac{1}{2} \log\left(\frac{1}{\mathrm{E}_t} - 1\right)$$

$$\exp(-2w_t) = \exp\left(-log\left(\frac{1}{\mathrm{E}_t} - 1\right)\right) = \exp\left(\log\left(\left(\frac{1}{\mathrm{E}_t} - 1\right)^{-1}\right)\right) = \left(\frac{1}{\mathrm{E}_t} - 1\right)^{-1} = \left(\frac{1 - \mathrm{E}_t}{\mathrm{E}_t}\right)^{-1}$$
$$= \frac{\mathrm{E}_t}{1 - \mathrm{E}_t}$$

Putting everything back:

$$\boxed{\mathrm{E}_{t+1} = \frac{\mathrm{E}_t}{\frac{\mathrm{E}_t}{1 - \mathrm{E}_t}(1 - E_t) + \mathrm{E}_t} = \frac{\mathrm{E}_t}{\mathrm{E}_t + \mathrm{E}_t} = \frac{1}{2}}$$