2. **Boosting**

We denote the weighted training error as the sum of all weighted errors over the sum of all weights.

$$\epsilon_t = \frac{W_l}{W_c + W_l} \tag{1}$$

We denote Z_t as the sum of the weights of each data point, that is

$$Z_t = W_c + W_l \tag{2}$$

Replacing ϵ_t in α with (1),

$$\alpha = \frac{1}{2} \cdot log\left(\frac{W_c}{W_l}\right) \tag{3}$$

(a) All notation defined in the problem set.

At each iteration t of T, all weights $w_i^{(t)}$ will be multiplied by $e^{-y_i \alpha_t h_t(x_i)}$ to give $w_i^{(t+1)}$.
The product $-y_i \cdot h_t(x_i)$ will give $+1$ when the prediction for the i-th data point matches its corresponding label, and -1 when the point is misclassified.
Hence, the weight of all correctly classified points will be multiplied by $e^{-\alpha}$, and the weight of all misclassified points will be multiplied by e^{α}.
That is:

$$W_l^{(t+1)} = W_l^{(t)} \cdot e^{\alpha}$$

and

$$W_c^{(t+1)} = W_c^{(t)} \cdot e^{-\alpha}$$

Using (2),

$$Z_{t+1} = W_l^{(t)} \cdot e^{\alpha} + W_c^{(t)} \cdot e^{-\alpha} \tag{4}$$

We call $z(\alpha)$ the function to compute Z_{t+1} from α (see equation (3)).

To minimize this function, we call $z'(\alpha)$ the derivative of $z(\alpha)$ with respect to α, set it to 0, and solve for α.

$$
\begin{aligned}
z'(\alpha) = \quad & W_l^{(t)} \cdot e^{\alpha} - W_c^{(t)} \cdot e^{-\alpha} = 0 \\
\iff \quad & W_l^{(t)} \cdot e^{\alpha} = W_c^{(t)} \cdot e^{-\alpha} \\
\iff \quad & \frac{e^{\alpha}}{e^{-\alpha}} = \frac{W_c^{(t)}}{W_l^{(t)}} \\
\iff \quad & 2\alpha = log\left(\frac{W_c^{(t)}}{W_l^{(t)}}\right)
\end{aligned}
$$

$$
\iff \quad \boxed{\alpha = \frac{1}{2} \cdot log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)}
$$

We find that $\alpha = \frac{1}{2} log \frac{1-\epsilon_t}{\epsilon_t}$ is optimal in the sense that it minimizes Z_{t+1}

(b) To show that Z_t is monotonically decreasing, we will start with the assumption that $Z_t < Z_{t+1}$, and prove that this is true.

$$
\begin{aligned}
& Z_{t+1} < Z_t \\
& \qquad \iff \quad W_l^{(t)} \cdot e^{\alpha} + W_c^{(t)} \cdot e^{-\alpha} < W_l^{(t)} + W_c^{(t)} \\
& \qquad \iff \quad W_l^{(t)} \cdot (e^{\alpha} - 1) + W_c^{(t)} \cdot (e^{-\alpha} - 1) < 0 \\
& \qquad \iff \quad \frac{e^{\alpha} - 1}{e^{-\alpha} - 1} < -\frac{W_c^{(t)}}{W_l^{(t)}} \\
& \qquad \iff \quad e^{\alpha} < \frac{W_c^{(t)}}{W_l^{(t)}} \\
& \qquad \iff \quad \alpha < log\left(\frac{W_c^{(t)}}{W_l^{(t)}}\right) \\
& \qquad \iff \quad \frac{1}{2} \cdot log\left(\frac{1-\epsilon_t}{\epsilon_t}\right) < log\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \\
& \qquad \iff \quad \boxed{\frac{1}{2} < 1}
\end{aligned}
$$

The statement holds, therefore, for all Z_t, Z_{t+1} is smaller. Hence, the sum of weights Z_t is monotonically decreasing as a function of t.