Q 1.1

As given from qs.

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$$

To find the training error ϵ_t we first go for the $t+1$ value from t i.e.

$$D_{t+1}(i) = D_1(i) \times \frac{e^{-y_i \alpha_1 h_1(x_i)}}{Z_1} \cdots \times \frac{e^{-y_i \alpha_T h_T(x_i)}}{Z_T}$$

$$= \frac{D_1(i) \exp\left(-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right)}{\prod_{t=1}^{T} Z_t}$$

or

$$\frac{D_1(i) \exp(-y_i f(x_i))}{\prod_{t=1}^{T} Z_t}.$$

Now we know that $H(x) = \text{sign}(f(x))$

if $H(x) \neq y$, then $yf(x) \leq 0$

which implies that $e^{-y f(x)} \geq 1$ or $= 1$. Thus,

$$\mathbb{1}\{H(x) \neq y\} \leq e^{-y f(x)}.$$

The weighted training error ε_{train}

$$Pr_{i \sim D_1}[H(x_i) \neq y_i] = \sum_{i=1}^{m} D_1(i) \mathbb{1}\{H(x_i) \neq y_i\}$$

or

$$\frac{1}{N} \sum_{j=1}^{N} \delta(H(x_j) \neq y_j) \leq \frac{1}{N} \sum_{j=1}^{N} \exp(-f(x_j) y_j)$$

<u>Note</u>: I have used notation diff. from that ~~of~~ of the assignment. The assignment has ~~give~~ used the superscript notation and here the subscript notation has been used.

1.2 So, ~~we~~ we know that

$$w_j^{(t+1)} Z_t = w_j^{(t)} \exp(-\alpha_t y_j h_t(x_j))$$

or $$w_j^{(t+1)} Z_t = \frac{1}{N} \sum_{j=1}^{N} w_j^{(t)} \exp(-\alpha_t y_j h_t(x_j))$$

or $$\boxed{Z_t = \frac{1}{N} \sum_{1}^{N} w_j^{(t)} \exp(-f(x_j) y_j)}$$

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

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

Q.1.3

(a) From the eqs.

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

$$= \sum_{i: y_i = h_t(x_i)} D_t(i) e^{-\alpha_t} + \sum_{i: y_i \neq h_t(x_i)} D_t(i) e^{\alpha_t}$$

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

$$= \prod_t \left[2\sqrt{\varepsilon_t(1-\varepsilon_t)}\right]$$

(b) Now, for (b) we can rewrite ε_t as $\frac{1}{2} - \gamma_t$. s.t. $\gamma_t > 0$.

To Prove:

$$Z_t \leq \exp(-2\gamma_t^2)$$

Let us replace value of $\varepsilon_t = \frac{1}{2} - \gamma_t$ in (1)

we get,

$$= e^{-\alpha_t}\left(\frac{1}{2} + \gamma_t\right) + e^{\alpha_t}\left(\frac{1}{2} - \gamma_t\right)$$

or $Z = \sqrt{1 - 4\gamma_t^2}$.

We know that (from previous proof) that the upper bound of the training error is $\prod_t Z_t$.
We also know that the training error drops rapidly when each weak classifier is assumed to have error bounded away from $\frac{1}{2}$.

We can see thus that the combined classifier will have training error at most

$$\left(\sqrt{1-4\gamma^2}\right)^T \leq e^{-2\gamma^2 T}.$$

or

$$Z_t \leq \exp\left(-2\sum_t^T \gamma_t^2\right).$$

(c) Now from #(b), we know that

$$\mathcal{E}_{train} \leq \exp\left(-2 \sum_{1}^{T} \gamma_t^2\right)$$

Now, as every classifier is better than random, we can just sum it up. and hence

$$\mathcal{E}_{train} \leq \exp\left(-2T\gamma^2\right).$$