PART 1

Q1 Error for each model:-

$$\varepsilon_i(x) = f(x) - h_i(x) \quad \text{---} (1)$$

(x : input & h_i : Bootstraped model)

Avg value for expected square error

$$E_{avg} = \frac{1}{m}\sum_{i=1}^{m} E(\varepsilon_i(x)^2) \quad \text{---} (2)$$

For the given case the aggregated model is defined as:

$$E_{agg}(x) = E\left[\left\{\frac{1}{m}\sum_{i=1}^{m} h_i(x) - f(x)\right\}^2\right]$$

$$= E\left[\frac{1}{m^2}\left[\sum_{i=1}^{m} h_i(x) - f(x)\right]^2\right]$$

$$= \frac{1}{m^2} E\left[\sum h_i(x) - f(x)\right]^2\right]$$

As $E[ax] = aE[x]$

From equation (1)

$$= \frac{1}{m^2} \sum E\left[[h_i(x) - f(x)]^2\right]$$

$$= \frac{1}{m}\left(\frac{1}{m}\sum E[\varepsilon_i(x)]^2\right]$$

Hence from equation (2)

$$\boxed{E_{agg}(x) = \frac{1}{m} \times E_{avg}}$$

Q2 Jensen's inequality

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leq \sum_{i=1}^{M} \lambda_i f(x_i) \quad \text{--- (1)}$$

Removing Assumption $E(\varepsilon_i(x)) \varepsilon_j(x)) = 0 \ \forall i \neq j$

i.e. $E(\varepsilon_i(x) \varepsilon_j(x)) \neq 0$

$$E_{avg} = \frac{1}{m} \sum_{i=1}^{m} E(\varepsilon_i(x)^2) \quad (\text{from Q 1})$$

so in $\lambda_i = \frac{1}{m}$ & $x_i > \varepsilon_i(x)$

let $f(x) = \lambda^2$, $\lambda = \sum_{i=1}^{m} \lambda_i x_i$

with these substitutions we can modify the Jensen's inequality equation as

$$\left(\sum \frac{1}{m} \varepsilon_i(x)\right)^2 \leq \sum_{i=1}^{m} \frac{1}{m} \varepsilon_i(x)^2$$

As, it is true for each value, it can be safely assumed that it is true for the average value as well. Hence taking avg on both sides we get.

$$E\left[\left\{\frac{1}{m} \sum_{i=1}^{m} \varepsilon_i(x)\right\}^2\right] \leq \frac{1}{M} \sum_{i=1}^{m} E(\varepsilon_i(x)^2)$$

Hence $\boxed{E_{agg} \leq E_{avg}}$

Q3 Hypothesis for boolean Classification problem

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad \text{--- (1)}$$

Weight for point i at step $t+1$

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times e^{-\alpha_t h_t(i) y(i)} \quad \text{--- (2)}$$

$D_t(i)$: Normalized weight of point i at time step t

$h_t(i)$: hypothesis at step t, for point i

α_t: Final voting power of hypothesis h_t

$y(i)$: True label for point i

Z_t: Normalization factor at step t --- (3)

At first step $D_1 = \frac{1}{N}$ for all points

Error: $E_t = \sum D_t(i)$ --- (4)

~~As~~ error is the sum of weights of misclassified points

$$D_{(t+1)}^{i} = \frac{D_t(i)}{Z_t} \times e^{-\alpha_t h_t(i) y(i)}$$

For multiple misclassified points

$$D_{t+1}(i) = D_t(i) \cdot \frac{e^{-\alpha_1 h_1(i) y(i)}}{Z_1} \times \frac{e^{-\alpha_2 h_2(i) y(i)}}{Z_2}$$

$$\cdots \times \frac{e^{-\alpha_t h_t y_i}}{Z_t}$$

Scanned with CamScanner

$$= \frac{1}{N} \frac{e^{-\sum_{j=1}^{t} \alpha_j h_j(i) y(i)}}{\prod_{j=1}^{t} Z_j}$$

where $f_t(i) = -\sum_{j=1}^{t} \alpha_j h_j(i)$

Now, total training error of $h(n)$

$$T_H = \frac{1}{N} \sum_{i = h(i) \neq y(i)}^{t} 1$$

Now sing $H(i) = \text{sign}(f(i))$

$$T_H = \frac{1}{N} \sum_{i = y(i) f(i) \leq 0}^{t} 1$$

because for misclassified points y_i & f_i would have opposite signs

hence $y_i f(i) \leq 0$

$$T_H \leq \frac{1}{N} \sum_{i}^{t} e^{-y(i) f(i)}$$

$$T_H \leq \left(\prod_t Z_t\right)\left(\sum_i D_{(t+1)}(i)\right)$$

D_{t+1} is a probability distribution

hence $\sum D_{(t+1)}(i) = 1$

$$T_H \leq \prod_t Z_t \quad \text{--- (6)}$$

$$Z_t = \sum_i D_t(i)\, e^{-\alpha_t h_t(i) y(i)}$$

$$= \sum_{i=h_t(i), y(i)} D_t(i) e^{-\alpha t} + \sum_{i=h_t(i) \neq y(i)} D_t(i) e^{\alpha t}$$

As for $h_t(i) = y(i)$ there sign will be equal so $h_t(i) \times y(i) = 1$

$$Z_t = e^{-\alpha t} \sum_{i=h_t(i)=y(i)} D_t(i) + e^{\alpha t} \sum_{i=h_t(i) \neq y(i)} D_t(i)$$

From eq (T)

$$Z_t = e^{-\alpha t}(1 - \varepsilon_t) + e^{\alpha t} \varepsilon_t$$

For minimising cost:

$$\varepsilon_t = \frac{1}{2} - \gamma_t \quad \text{[Given in question]}$$

$$Z_t = 2\sqrt{\left(\frac{1}{2} - \gamma_t\right)\left(\frac{1}{2} + \gamma_t\right)}$$

$$Z_t = \sqrt{1 - 4\gamma_t^2}$$

Since $1 + x \leq e^x \quad \forall x \in R$

$$Z_t \leq \sqrt{e^{-4\gamma_t^2}}$$

puting Z_t in eq 6

$$T_H \leq \prod_{t=1}^{T} e^{-2\gamma_t^2}$$

$$\boxed{T_H \leq e^{-2\sum_{t=1}^{T} \gamma_t^2}}$$