Question 1:

Solution:

Part 1 - Question :1

Solution :

Given → $\epsilon_i(x) = f(x) - h_i(x) \Rightarrow (1)$

(2) → $E(\epsilon_i(x)^2) = E[(f(x) - h_i(x))^2]$

(3) → $E_{avg} = \frac{1}{M} \sum_{i=1}^{M} E(\epsilon_i(x)^2)$

(4) → $h_{agg}(x) = \frac{1}{M} \sum_{i=1}^{M} h_i(x)$

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

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

Proof :

$E_{agg} = E\left[\left\{\frac{1}{M} \sum_{i=1}^{M} h_i(x) - f(x)\right\}^2\right]$ [From (5)]

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

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

[Since $h_i(x) - f(x) = (-1)(f(x) - h_i(x))$]

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

[Since $E[aX] = aE[X]$]

From ①

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

$$= \frac{1}{M} \cdot \underbrace{\frac{1}{M} E\left[\left\{\sum_{i=1}^{M} \epsilon_i(x)\right\}^2\right]}_{\text{From ③}}$$

$$= \frac{1}{M} E_{avg}$$

Hence proved

$$\boxed{E_{agg} = \frac{1}{M} E_{avg}}$$

Question 2:

Solution:

Part 1 - Question : 2

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

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

Assuming $g(x) = \sum x_i$

① $\Rightarrow$

$E_{agg} = E\left[\left\{\frac{1}{m} g(h_i(x) - f(x))\right\}^2\right] \rightarrow ③$

$E_{avg} = \left[\frac{1}{m} g\left(E[(f(x) - h_i(x))^2]\right)\right) \rightarrow ④$.

Jensen's inequality :- $f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i) \rightarrow ⑤$

Assuming $\lambda = 1/m$ and applying ⑤ on ③ & ④ $[E[g(x)] \geq g(E[x])]$

We can conclude that

$E_{agg} \leq E_{avg}$.

Question 3:

Solution:

Part 1 - Question: 3

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

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

Error for adaboost process, E_t, can be written wrt D_t as

$$E_t = \sum_{i\, h_t(i) \neq y(i)} D_t(i) \rightarrow ③$$

Since both $h_t(i)$ & $y(i)$ are in $\{-1, 1\}$ we can expand ② as:

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

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

$$= \frac{1}{N} e^{-y(i) f_t(i)} \rightarrow ④$$

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

The total training error of $H(x)$ can be written as:

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

[This is the average of the misclassified points]

We know $H(i) = \text{Sign}(f(i)) \longrightarrow (5)$

Hence, we can conclude

$$T_H = \frac{1}{N} \sum_{i = h(i) \neq y_{(i)}} 1 = \frac{1}{N} \sum_{i = y(i) f(i) \leq 0} 1$$

Since misclassified points have opposite signs.

Because we have $e^{-z} \geq 1$ when $z \leq 0$, we can write

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

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

[From (4)] $\leq \frac{1}{N} \left(N \prod_t Z_t\right) \sum_i D_{t+1}(i)$

$$\leq \frac{1}{\cancel{N}} \left(\cancel{N} \prod_t Z_t\right) \times 1$$

[Since D_{t+1} is a probability distribution. So its sum is 1]

$$T_H \leq \prod_t Z_t \longrightarrow (6)$$

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

[From (5)] $= \sum_{i=h_t(i)=y(i)} D_t(i)\, e^{-\alpha_t} + \sum_{i=h_t(i)\neq y(i)} D_t(i)\, e^{\alpha_t}$

Minimizing error T_H,
we find $\alpha_t = \frac{1}{2} \log \frac{1-E_t}{E_t} \longrightarrow (7)$

Putting (7) in the above equation:

$$Z_t = 2\sqrt{E_t(1-E_t)} \longrightarrow (8)$$

We are given:

$$E_t = \frac{1}{2} - \gamma_t \longrightarrow (9)$$

Putting (9) in (8):

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

$$= \cancel{2}\sqrt{\frac{(1-4\gamma_t^2)}{\cancel{4}}}$$

$$= \sqrt{1-4\gamma_t^2} \longrightarrow (10)$$

It is known that $1+x \leq e^{x} \; \forall \; x \in R$

$$\therefore \; 1-4\gamma_t^2 \leq e^{-4\gamma_t^2} \rightarrow (11)$$

Putting (11) in (10)

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

$$\Rightarrow Z_t \leq e^{-2\gamma_t^2} \rightarrow (12)$$

Putting (12) in (6) :

$$T_H \leq \prod_t Z_t$$

$$T_H \leq \prod_t e^{-2\gamma_t^2}$$

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

[Solution]