SXC220127
SXM210368

Assignment 3. PART 1

1) We need to prove that

$$E_{agg} = \frac{1}{m} E_{avg}$$

Given from the question:

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

also given that

$$E_{avg} = \frac{1}{m}\sum_{i=1}^{m} E(\varepsilon_i(x)^2) \longrightarrow (B)$$

Error for each of the model is given as:

$$\varepsilon_i(x) = f(x) - h_i(x) \rightarrow (C)$$

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

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

From eq (C) we know that

$$\varepsilon_i(x) = f(x) - h_i(x)$$

Substituting (C) in (D)

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

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

The above (E) can be also written as:

$$= \frac{1}{m}\underbrace{\frac{1}{m}\sum_{i=1}^{M} E\left[\varepsilon_i(x)^2\right]}_{\rightarrow E_{avg} \text{ from } (B)}$$

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

(2) From the question we need to prove that

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

It is given that:

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

Proof: If $f \rightarrow$ complex function on (a, b)

$x \rightarrow$ random variable.

then,

$$f(E(x)) \leq E(f(x)) \rightarrow (i)$$

Now,

$$E(f(x)) = \lambda(x_i) f(x_i) + \sum_{i=2}^{M} \lambda(x_i) f(x_i)$$

$$= \lambda(x_1) f(x_1) + (1-\lambda(x_1)) \sum_{i=2}^{M} \frac{\lambda(x_i) f(x_i)}{1-\lambda(x_1)}$$

$$\leq \lambda(x_1) f(x_1) + (1-\lambda(x_1)) f\left(\sum_{i=2}^{M} \lambda(x_i)(x_i)^{(1-p(x))}\right)$$

$$\leq f\left(\lambda(x_1) x_1 + 1 - p(x_1)\left(\sum_{i=2}^{M} \frac{\lambda(x_i) x_i}{1-\lambda(x_1)}\right)\right)$$

$$\leq f(E(x))$$

From the above (i) we can conclude that

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

③ From the question, given that:

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

At step $t+1$ is given by.

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

Also error at time t. $\epsilon_t = \sum D_t(i)$, $h(i) \neq y(i)$

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

Sum of weights corresponding to all i points that are not misclassified at error time t.

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

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

$$= \frac{1}{N} \sum_{i\, h(i) \neq y(i)} 1 \qquad (H(i) = \text{sign}(f(i)))$$

Average of misclassified points

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

$y(i)$ & $f(i)$ will be opposite sign

that is $\Rightarrow y(i) f(i) \leq 0$

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

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

$$\leq \frac{1}{N} \times N \prod_t Z_t \sum_i D_{t+1}(i)$$

$\Sigma_i D_{t+1}(i) = 1$, as D_{t+1} is a probability distribution

$$T_H \leq \prod_t Z_t$$

$$Z_t = \sum_{i\, h_t(i) = y_t(i)} D_t(i) e^{-\alpha t} + \sum_{i\, h_t(i) \neq y_t(i)} D_t(i) e^{\alpha - t} \rightarrow (1)$$

In (1), $h_t(i) y(i) = 1$ iff $h_t(i) = y(i)$

$h_t(i) y(i) = -1$ iff $h_t(i) \neq y(i)$

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

As we minimize the error T_H, α_t will be $\frac{1}{\alpha} \ln \frac{1-\varepsilon_t}{\varepsilon_t}$

$$Z_t = 2\sqrt{\varepsilon(1-\varepsilon_t)} \quad \because \varepsilon_t = \frac{1}{2} - \gamma_t$$

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

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

As, $1 + x \leq e^x$

$$1 - 4\gamma_t^2 \leq e^{-4\gamma_t^2}$$

$$Z_t = e^{-2\gamma_t^2}$$

$$T_H \leq \prod_t Z_t$$

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

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