CS 6375
ASSIGNMENT 3

Names of students in your group:
Aaryan Singh Chauhan (axc230019)
Nikunj Gohil (ndg220000)

Number of free late days used: 0

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

ASSIGNMENT 3

PART 1

1. Error of each model : $\epsilon_i(x) = f(x) - h_i(x)$

Expected squared error :

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

Average value of Expected squared error :

$$E_{avg} = \frac{1}{M} \sum E(\epsilon_i(x)^2)$$

Aggregrate model : $h_{agg}(x) = \frac{1}{M} \sum h_i(x)$

Error of Aggregrate model : $E_{agg}(x) = E\left[\left\{\frac{1}{M} \sum h_i(x) - f(x)\right\}^2\right]$

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

P.T : $E_{agg} = (1/M)\, E_{avg}$

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

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

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

$\because \left(E(ax + b) = aE(x) + b\right).$

No. Date

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

$$(\because (-(h_i(x) - f(x))^2 = (f(x) - h_i(x))^2 = \epsilon_i(x)^2)$$

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

$$(\because E\left[\left\{\sum \epsilon_i(x)\right\}^2\right] = E\left[\epsilon_1^2(x) + \epsilon_2^2(x) + \dots \right.$$
$$\dots + \epsilon_m^2(x) + 2[\epsilon_1(x)\epsilon_2(x) + \epsilon_1(x)\epsilon_3(x) + \dots$$
$$\left.\dots + \epsilon_{m-1}(x)\epsilon_m(x)]\right]$$

Given assumption : $\epsilon_i(x)\epsilon_j(x) = 0 \Rightarrow$

$$E\left[\left\{\sum \epsilon_i(x)\right\}^2\right] = E[\epsilon_1^2(x) + \dots + \epsilon_m^2(x)]$$
$$= E[\epsilon_1^2(x)] + \dots + E[\epsilon_m^2(x)]$$

$$E_{agg} = \frac{1}{m} E_{avg} \qquad (\because E_{avg} = \frac{1}{m} \sum E[\epsilon_i^2(x)]$$

2. Given Jensen's inequality :

$$f\left(\sum \lambda_i x_i\right) \leq \sum \lambda_i f(x_i)$$

Let us define a function $f(x) = E[x^2]$.

$\therefore$ when $x = \sum_{i=1}^{m} \frac{1}{m} \epsilon_i(x)$,

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

$$(\because \sum a x_i = a \sum x_i \ ; \ E_{agg} = E\left[\left\{\frac{1}{m} \sum \epsilon_i(x)\right\}^2\right])$$

Using the above inequality,

$$E_{agg} \leq \sum_{i=1}^{m} \frac{1}{m} E[\epsilon_i^2(x)]$$

Date

MARUTI

$$E_{agg} \le \frac{1}{m} \sum_{i=1}^{m} E[\epsilon_i^2(x)]$$

$$(\because \sum a x_i = a \sum x_i).$$

$$\therefore E_{agg} \le E_{avg} \qquad \left(\because E_{avg} = \frac{1}{m} \sum E[\epsilon_i^2(x)]\right)$$

3. Given, $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times e^{-\alpha_t h_t(i) y_t(i)}$

$$= \frac{e^{-\alpha_t h_t(i) y_t(i)}}{Z_t} \cdot D_t(i)$$

$$= \left(\frac{e^{-\alpha_t h_t(i) y_t(i)}}{Z_t}\right)\left(\frac{e^{-\alpha_{t-1} h_{t-1}(i) y_{t-1}(i)}}{Z_{t-1}} \cdot D_{t-1}(i)\right).$$

and if we continue expanding,

$$D_{t+1}(i) = \left(\frac{e^{-\alpha_t h_t(i) y_t(i)}}{Z_t}\right) \cdots \left(\frac{e^{-\alpha_1 h_1(i) y_1(i)}}{Z_1} D_1(i)\right)$$

$$D_{t+1}(i) = \frac{e^{-y_i(i) \sum \alpha_t h_t(i)}}{\prod_{t=1}^{t} Z_t} \left(\frac{1}{N}\right) \qquad (\because D_1 = 1/N). \qquad — (1).$$

Let $f(i) = \sum \alpha_t h_t(i)$

Total training error is the average of misclassified points,

$$\therefore E_{train} = \frac{1}{N} \sum_i \begin{cases} 1 & \text{if } y(i) \neq h(i) \\ 0 & \text{otherwise.} \end{cases}$$

$$= \frac{1}{N} \sum_i \begin{cases} 1 & \text{if } y(i) f(i) \le 0 \\ 0 & \text{else} \end{cases}$$

$$= \frac{1}{N} \sum e^{-y(i) f(i)}$$

$$\left(\because \text{if } z \le 0 \Rightarrow e^{-z} \ge 1 \atop z > 0 \Rightarrow e^{-z} < 1\right)$$

Date

From ①, $D_{t+1}(i) = \frac{1}{N} \frac{e^{-y_i f(i)}}{\prod_t Z_t}$

$\therefore \left(\frac{1}{N} e^{-y_i f(i)}\right) = D_{t+1}(i) \prod_t Z_t$.

$\therefore E_{train} = \sum_i D_{t+1}(i) \prod_t Z_t$

$E_{train} = \prod_t Z_t$ ($\because D_{t+1}$ is distribution, $\sum D_{t+1} = 1$).

Z_t is the normalizing factor,

$\therefore Z_t = \sum_t D_t(i) \cdot \begin{cases} e^{-\alpha t} & \text{if } h_t(i) = y_i \\ e^{\alpha t} & \text{if } h_t(i) \neq y_i \end{cases}$

$= \sum D_t(i) e^{-\alpha t} + \sum D_t(i) e^{\alpha t}$

$= e^{-\alpha t} \sum D_t(i) + e^{\alpha t} \sum D_t(i)$

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

($\because \epsilon_t = \sum_i D_t(i) \quad h_t(i) \neq y(i)$ and similarly).

$Z_t = 2\sqrt{\epsilon_t(1-\epsilon_t)}$ ($\because \alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$ to minimize the error).

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

($\because \epsilon_t = \frac{1}{2} - \gamma_t$).

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

$\therefore Z_t \leq e^{-2\gamma_t^2}$ ($\because 1 + x \leq e^x$).

$\therefore E_{train} = \prod_t Z_t \leq \prod_t e^{-2\gamma_t^2} \Rightarrow E_{train} \leq e^{-2\sum_{t=1}^{T} \gamma_t^2}$