

CS 4375

ASSIGNMENT _____3_____

Names of students in your group:

John Kenney

Mohammad Syed

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.

Part 1

1. Assume

1. $E(\varepsilon_i(x)) = 0$ for all i

2. $E(\varepsilon_i(x)\varepsilon_j(x)) = 0$ for all $i \neq j$

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

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

Since

$$\left(\sum_{i=1}^M a_i\right)^2 = \sum_{i=1}^M \sum_{j=1}^M a_i a_j$$

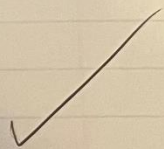
$$E\left[\frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \varepsilon_i(x) \varepsilon_j(x)\right] = \frac{1}{M^2} \sum_{i=1}^M E(\varepsilon_i(x) \varepsilon_i(x))$$

$$\frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M E[\varepsilon_i(x) \varepsilon_j(x)] = 11$$

Since

2. Every value of j except when $j=i$ will be zero

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



2.

$$f(x) = x^2$$

$$\left(\sum_{i=1}^M \lambda_i x_i\right)^2 \leq \sum_{i=1}^M \lambda_i x_i^2$$

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

Since

Squared use as $f(x)$

So then

Jensen inequality can be used

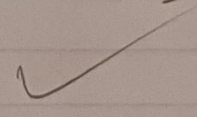
$$E_{agg} = E\left[f\left(\frac{1}{M} \sum_{i=1}^M \varepsilon_i(x)\right)\right]$$

by Jensen inequality

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

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

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



3) Error E_t ~~Adaboost~~ Adaboost process can be measured w/ respect to D_t as

$$E_t = \sum D_t(i)$$

error at time t is the sum of weights corresponding to all points i which are mis-classified $\rightarrow h_t(i) \neq y_i$

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

since $h_t(i)$ & $y(i)$ are both in $\{1, -1\}$, this recurrence gives

$$\begin{aligned} D_{t+1}(i) &= D_t(i) \cdot \frac{e^{-\alpha_1 h_1(i) y(i)}}{Z_1} \cdot \frac{e^{-\alpha_2 h_2(i) y(i)}}{Z_2} \cdots \frac{e^{-\alpha_t h_t(i) y(i)}}{Z_t} \\ &= \frac{1}{N} \frac{e^{-\sum_{j=1}^t \alpha_j h_j(i) y(i)}}{\prod_{j=1}^t Z_j} \end{aligned}$$

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

Total training error of $H(x)$

$$T_H = \frac{1}{N} \sum \dots \text{ie; avg of mis-classified points}$$

Since $H(i) = \text{sign}(F(i))$

$$\therefore T_H = \frac{1}{N} \sum_i 1_{y(i) \neq F(i)} \leq 0$$

b/c for mis-classified $y(i) \neq F(i)$ would have opposite signs, hence $y(i)F(i) \leq 0$

$$\text{Then } T_H = \frac{1}{N} \sum_i 1_{y(i)F(i) \leq 0} \leq \frac{1}{N} \sum_i e^{-y(i)F(i)}$$

$$\left(\begin{array}{l} \text{b/c } e^{-z} \geq 1 \\ \text{when } z \leq 0 \end{array} \right)$$

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

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

= 1 b/c $(D_t + 1)$ is a probability distribution

to ~~reach~~

$$\Rightarrow T_H \leq \prod_t Z_t$$

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

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

b/c for ~~any~~ $h_t(i) = y(i)$

$h_t(i) y(i) = 1$ as both would be 1 or -1

$$\Rightarrow Z_t = e^{-\alpha_t} (1 - \epsilon_t) + e^{\alpha_t} \epsilon_t$$

Now to minimize error ~~with~~ T_H , it comes out to be $\frac{1}{2} \frac{1 - t_t}{t_t}$

$$\text{So } Z_t = 2\sqrt{t_t(1-t_t)}$$

$$\text{now, } t_t = 1/2 - Y_t \quad \text{given}$$

$$\text{the } Z_t = 2\sqrt{(1/2 - Y_t)(1/2 + Y_t)} = \sqrt{1 - 4Y_t^2}$$

$$\text{since, } 1+x \leq e^x \quad \forall x \in \mathbb{R}$$
$$\Rightarrow 1 - 4Y_t^2 \leq e^{-4Y_t^2}$$

$$\therefore Z_t \leq \sqrt{e^{-4Y_t^2}} = e^{-2Y_t^2}$$

Putting Z_t in gives

$$T_H \leq \prod_t Z_t$$

$$\Rightarrow T_H \leq \prod_t e^{-2Y_t^2}$$

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