# 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. $E(\mathcal{E}_i(x)) = 0$ for all $i$
   2. $E(\mathcal{E}_i(x)\mathcal{E}_j(x)) = 0$ for all $i \neq j$

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

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

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

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

$$\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}E[\mathcal{E}_i(x)\mathcal{E}_j(x)] = \quad '' \qquad ''$$

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

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

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

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

since squared use as $f(x)$

$$E_{agg} = E\left[f\left(\frac{1}{M}\sum_{i=1}^{M}\mathcal{E}_i(x)\right)\right] \text{ by jenson inequality}$$

so then Jenson inequality can be used

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

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

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

3) Error $\epsilon_t$ [redacted] Adaboost process can be measured w/ respect to $D_t$ as

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

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 recurance gives

$$D_{t+1}(i) = D_1(i) \cdot \frac{e^{-\alpha_1 h_1(i)\, y(i)}}{Z_1} \cdot \frac{e^{-\alpha_2 h_2(i) y(i)}}{Z_2} \cdots \frac{e^{-t}}{Z_t}$$

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

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

Total training ~~err~~ error of $H(x)$

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

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

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

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

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

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

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

$\Rightarrow \; T_H \leq \left( \prod_t Z_t \right) \underbrace{\left( \sum D_{t+1}(i) \right)}_{= 1 \text{ b/c } D_{t+1} \text{ is a probability distribution}}$

$\Rightarrow \; T_H \leq \prod Z_t$

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

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

b/c for ~~[illegible]~~ $h_{(t)}(i) = y(i)$
$h_t(i) \, y(i) = 1$ 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 ~~..~~ $T_H$, $\alpha_t$ comes out to be $\frac{1}{\alpha} \quad \frac{1-\epsilon_t}{\epsilon_t}$

so $\quad Z_t = 2\sqrt{\epsilon_t(1-\epsilon_t)}$

now, $\quad \epsilon_t = \frac{1}{2} - \gamma_t \quad$ given

the $Z_t = 2\sqrt{(\frac{1}{2} - \gamma_t)(\frac{1}{2} + \gamma_t)} = \sqrt{1 - 4\gamma_t^2}$

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

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

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

Putting $Z_t$ in gives

$$T_H \leq \prod_t Z_t$$

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

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