PART ONE

Terisha Prax

1)

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

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

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

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

since we assume errors are uncorrelated $E(\varepsilon_i \cdot \varepsilon_j) = 0$ for all $i \neq j$, all that is remaining is $E(\varepsilon_i \cdot \varepsilon_j)$ where $i=j$ so we can write it as $E(\varepsilon_i^2)$

$$= \frac{1}{M} \cdot \underbrace{\frac{1}{M} \cdot E\left(\sum_{i=1}^{M} \varepsilon_i(x)^2\right)}_{E_{avg}(x)}$$

$$E_{agg}(x) = \frac{1}{M} \cdot E_{avg}(x)$$

2)

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

if we define $f(x)$ as taking the expectation & λ_i as $\frac{1}{M^2}$ & $x_i = \varepsilon_i(x)$ then we can rewrite

$$E_{agg}(x) = f\left(\sum_{i=1}^{M}\lambda_i x_i\right) \quad \& \quad E_{avg}(x) = \sum_{i=1}^{M}\lambda_i f(x_i)$$ → note $\sum_{i=1}^{M}\frac{1}{M^2} = \frac{M}{M^2} = \frac{1}{M}$

so then by Jensen's inequality, $E_{agg} \leq E_{avg}$

3) We are given that $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times e^{-\alpha_t h_t(i) y(i)}$. Let t go from $1:T$ - then to get the overall training error we would look at the the normalized weight after iteration T. If we write $\alpha_t h_t(i) = f(x)$ then we can say:

$$D_{T+1}(i) = D_1(i) \times \frac{e^{-f_1(x_i)\cdot y(i)}}{Z_1} \times \frac{e^{-f_2(x_i)\cdot y(i)}}{Z_2} \times \ldots \times \frac{e^{-f_T(x_i)\cdot y}}{Z_T}$$

we are given that $D_1 = \frac{1}{N}$ and we know $e^a \times e^b = e^{a+b}$.

$$D_{T+1}(i) = \frac{\frac{1}{N} \times e^{-y_i\left(\sum_t f_t(x_i)\right)}}{\prod_t Z_t}.$$

Then we can calc the training error of H, which we define as the last classifier.

$$\varepsilon(H) = \frac{1}{N}\sum_i \begin{cases} 1 & \text{if } y_i \neq H(x_i) \\ 0 & \text{o.w.} \end{cases}$$ per def of misclassific. error

We are given that $H(x_i) = \text{sign}(f(x))$. So that means $y_i \neq \text{sign}(f(x))$ so y_i must be either -1 or $+1$ (opp of sign)

so $$\varepsilon(H) = \frac{1}{N}\sum_i \begin{cases} 1 & \text{if } y_i \cdot \text{sign}(f(x)) \leq 0 \\ 0 & \text{o.w.} \end{cases}$$

You can then say that $\varepsilon(H) \leq \frac{1}{N}\sum_i e^{-y_i f(x_i)}$ since $e^{-x} \geq 1$ if $x \leq 0$ & we said $y_i \cdot \text{sign}(f(x)) \leq 0$ but here we have $-y_i \cdot f(x_i)$ so we have the inequality.

→

we can then substitute b/c we know $D_{T+1}(i) = \frac{1}{N} \frac{e^{-y_i \sum_t f_t(x_i)}}{\prod_t Z_t}$

so $\quad \varepsilon(H) \leq \sum_i D_{T+1}(i) \times \prod_t Z_t$

$D_{T+1}(i)$ is a distribution of normalized weights so we can say that the training error $\leq \prod_t Z_t$

so then the question becomes, what is Z_t?

we know it is the normalization factor so we can say

$$Z_t = \sum_i D_t(i) \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases}$$

$$= \underset{@\text{ all obs}}{\text{normalized wt}} \times \begin{cases} e^{-\alpha_t} & \text{if hypothesis = true label} \\ e^{\alpha_t} & \text{if hypothesis} \neq \text{true label} \end{cases}$$

$$= \sum_{\substack{\text{all } i \\ \text{where} \\ h_t(x_i) = y_i}} D_t(i) \times \underbrace{e^{-\alpha_t}}_{\text{no } i} + \sum_{\substack{\text{all } i \\ \text{where} \\ h_t(x_i) \neq y_i}} D_t(i) \times \underbrace{e^{\alpha_t}}_{\text{no } i}$$

$$= e^{-\alpha_t} \underbrace{\sum_{\substack{\forall i \text{ where} \\ h_t(x_i) = y_i}} D_t(i)}_{\text{all the right/ correct wts}} + e^{\alpha_t} \underbrace{\sum_{\substack{\forall i \text{ where} \\ h_t(x_i) \neq y_i}} D_t(i)}_{\text{incorrect wts / error}}$$

$$= e^{-\alpha_t}(1 - \varepsilon_t) + e^{\alpha_t}(\varepsilon_t)$$

double check

* if we let $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ then we can rewrite

$$Z_t = e^{-\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \cdot \frac{1}{2}} (1-\varepsilon_t) + e^{\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \cdot \frac{1}{2}} (\varepsilon_t)$$

$$= 2(\varepsilon_t(1-\varepsilon_t)^{\frac{1}{2}})$$

if we plug in our given $\varepsilon_t = \frac{1}{2} - \gamma_t$ then

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

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

& since we know $1 + x \leq e^x \ \forall x$

$Z_t \leq e^{-2\gamma_t^2}$ so training error $\leq \prod_t e^{-2\gamma_t^2} \rightarrow \boxed{\text{t.e.} \leq e^{-2\sum_t \gamma_t^2}}$