Q1. $$E_{avg} = \frac{1}{M}\sum_{i=1}^{M} E\left(\varepsilon_i(x)^2\right) \quad \text{——— (1)}$$

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

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

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

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

[As we know that $E(ax) = a\,E(x)$]

$$= \frac{1}{M^2} E\left[\left(\varepsilon_1(x) + \varepsilon_2(x) + \varepsilon_3(x) + \cdots + \varepsilon_M(x)\right)^2\right]$$

$$= \frac{1}{M^2} E\left[(\varepsilon_1(x))^2 + (\varepsilon_2(x))^2 + \cdots + (\varepsilon_M(x))^2 + 2\cdot\varepsilon_1(x)\,\varepsilon_2(x) + \cdots\right]$$

$$= \frac{1}{M^2}\left[E\left[\sum_{i=1}^{M} \varepsilon_i(x)^2\right] + 2E\left[\underbrace{\sum_{i=0}^{M}\sum_{j=1}^{M}\left(\varepsilon_i(x)\cdot\varepsilon_j(x)\right)}_{=0}\right]\right]$$

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

[As all ~~are~~ errors are uncorrelated and $E\left(\varepsilon_i(x)\,\varepsilon_j(x)\right) = 0$ for all $i \neq j$)

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

[from (1)]

Hence,

$$\boxed{E_{agg}(x) = \frac{1}{M} \cdot E_{avg}}$$ [from ① and ②]

Hence Proved.

Q2.

Jensen's inequality states that for any convex function f :

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

[assuming that each of the errors are not uncorrelated.]

To prove : $E_{agg} \leq E_{avg}$

Ans. →

Now, as the errors are not uncorrelated i.e.

$$E\left(\varepsilon_i(x) \cdot \varepsilon_j(x)\right) \neq 0 \quad \text{for all } i \neq j$$

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

and

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

comparing with Eavg with RHS of Jensen's inequality, according to the given equation

$$\lambda_i = \frac{1}{M}, \quad \cancel{f(x_i) = \sum_{i=1}^{M} E[\varepsilon_i(x)^2]}$$

$$f(x_i) = E[\varepsilon_i(x)^2]$$

and we have,

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

i.e. we get from Jensen's inequality as:

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

we need to prove that,

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

Because we have already proved from result ① that

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

we know that

$$\frac{(a+b)^2}{2} \leq a^2+b^2$$

(from Cauchy-Schwarz inequality)

$$\frac{(\varepsilon_1(x) + \varepsilon_2(x) + \cdots + \varepsilon_M(x))^2}{M} \leq \varepsilon_1(x)^2 + \varepsilon_2(x)^2 + \cdots + \varepsilon_M(x)^2$$

$$E\left[\frac{1}{M}\left(\sum_{i=1}^{M} \varepsilon_i(x)\right)^2\right] \leq E\left[\varepsilon_1(x)^2 + \varepsilon_2(x)^2 + \cdots + \varepsilon_M(x)^2\right]$$

dividing each side by M,

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

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

We know that, from the question 1,

$$E_{agg}(x) = E\left[\left\{\frac{1}{M}\sum_{i=1}^{M} \varepsilon_i(x)\right\}^2\right] \quad \text{——— (3)}$$

and

$$E_{avg} = \frac{1}{M} \cdot \sum_{i=1}^{M} E(\varepsilon_i(x)^2) \quad \text{——— (4)}$$

from (2), (3) and (4), we get

$$\boxed{E_{agg}(x) \leq E_{avg}}$$

Hence Proved.

Q3. Final Hypothesis for a Boolean classification problem at the end of T iterations is given as:

$$H(X) = \text{sign}\left(\sum_{t=1}^{T} a_t h_t(X)\right) \quad \text{——— (1)}$$

i.e. Final hypotheses is the weighted hypothesis generated at the end of each individual step. The weight for the point i at step t+1 is given by :

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times e^{-\alpha_t h_t(x_i)\, y(i)} \quad \text{——— (2)}$$

at step 1, the points have equal weight

$$D_1 = \frac{1}{N}. \quad \text{——— (3)}$$

total error at each step is:

$$\varepsilon_t = \frac{1}{2} - \gamma_t \quad \text{——— (4)}$$

We have,

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t\, y_i\, h_t(x_i))}{Z_t}$$

$$= \frac{D_t(i) \cdot e^{-\alpha_t \cdot y(i) \cdot h_t(x_i)}}{Z_t}. \quad \text{from (2)}$$

Since, $y(i)$ and $h_t(x_i)$ are both in $\{-1, +1\}$ unwrapping this recurrence,

$$D_{T+1}(i) = D_1(i) \cdot \frac{e^{-\alpha_1 y_i h_1(x_i)}}{Z_1} \cdots \frac{e^{-\alpha_T y_i h_T(x_i)}}{Z_T}$$

$$= \frac{1}{N} \cdot \frac{e^{(-(y_i)\sum_t \alpha_t h_t(x_i))}}{\prod_t Z_t}$$

$$= \frac{1}{N} \cdot \frac{\exp(-y_i f(x_i))}{\prod_t Z_t} \quad \text{——— (5)}$$

Now,

$$\text{training error }(H) = \frac{1}{N} \sum_i \begin{cases} 1 & \text{if } y_i \neq H(x_i) \\ 0 & \text{else} \end{cases}$$

(By the definition of training error)

$$= \frac{1}{N} \sum_i \begin{cases} 1 & \text{if } y_i f(x_i) \leq 0 \\ 0 & \text{else} \end{cases} \quad \Big| \text{ since } H(x) = \text{sign } f(x) \text{ \& } y_i \in \{-1, +1\}$$

$$\leq \frac{1}{N} \sum_i \exp(-y_i f(x_i)) \quad \Big| \text{ as } e^{-z} \geq 1 \text{ if } z \leq 0$$

$$= \sum_i D_{T+1}(i) \cdot \prod_t Z_t \quad \Big| \text{ from step (5)}$$

$$= \prod_t Z_t \quad \Big| \text{ since} \quad \text{——— (6)}$$

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

$$= \sum_{i: h_t(x_i) = y_i} D_t(i) \cdot e^{-\alpha t} + \sum_{i: h_t(x_i) \neq y_i} D_t(i) \cdot e^{\alpha t}$$

$$= e^{-\alpha t} \sum_{i: h_t(x_i) = y_i} D_t(i) + e^{\alpha t} \sum_{i: h_t(x_i) \neq y_i} D_t(i)$$

$$= e^{-\alpha t}(1 - \epsilon_t) + e^{+\alpha t} \cdot \epsilon_t \quad \text{(by the definition of } \epsilon_t)$$

$$= 2\sqrt{\epsilon_t(1-\epsilon_t)}$$ (as α_t was chosen to minimize the expression)

$$= \sqrt{1-4\gamma_t^2}$$ plugging in $\epsilon_t = \frac{1}{2} - \gamma_t$

$$\leq e^{-2\gamma_t^2}$$ using $1+x \leq e^x$ for all real x.

——————— 7

from 6 and 7,

training error $= \prod_t Z_t$

& $\quad Z_t \leq e^{-2\gamma_t^2}$

Hence, this gives the claimed upper bound on the training error of H.

Hence, at the end of T steps, overall training error will be less than or equal to ~~the~~ i.e. it will be bounded by:

$$\boxed{\exp\left(-2\cdot\sum_{t=1}^{T}\gamma_t^2\right)}$$ [taking t=1 to t=T]

Hence proved.