(f.k)
Order(order-id, customer-id, order-date, Total-order-amount,
Service-method, order-time)

(f.k)
order-items (order-id, Item-id, Item-name, Quantity-ordered,
price)

(fk)
Customizations(Item-id, cheese, sauce, size, crust,
cut, Bake, seasoning)

(fk)
Toppings (Item-id, toppings-meat, toppings-non meat,
option selected)

(f.k)
payments(payment-id, order-id, coupon code,
payment-date, payment status, payment-status
payment-type, Total-amount)

```
{
order-id : "1",
customer-id : "1234,"
order date : "15-04-2021"
Order time : "10:30:45 AM",
Service method : "Online",
Total order amount : "25",
Order-item ;
{
Item-id : "PJ123",
Item-name : "PIZZA",
Customizations :
{
Cheese : "Normal",
Sauce : "Hearty Marinara sauce",
Size : "Large",
Crust : "hand tossed".
Bake ; "Well Done",
Seasoning : "Garlic seasoned crust",
Cut : "Pie cut",
Toppings :
{
Toppings-meat ; ["Premium chicken", "Bacon"]
option-selected : ["Normal", "Extra"]
Toppings-non meat : ["Jalapeno peppers", Pineapple"]
option-selected : ["Normal", "Extra"]
}
```

}
Quantity - ordered : "2",
Price : "15"
}
payment:
{
payment - id : "123456",
Coupon code : "ABCDEFG",
payment - date : "15-10-2021",
Payment - status : "Paid",
payment - type : "zelle",
Total amount : "20"
}
}

Question 1 :-

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

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

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

→ we are also given assumptions that Each of the errors have a 0 mean

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

and errors are uncorrelated

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

From (1),

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

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

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

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

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

According to our question,

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

Hence, $E_{agg}(x) = \frac{1}{M} \times E_{avg}$

Question 2:-

Jensen's inequality states that for any convex function f

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

$$E(\epsilon_i(x)\,\epsilon_j(x)) \neq 0, \quad \forall\, i \neq j$$

and we have to assume that each of the errors generated from different models using bootstrap samples are correlated.

Ø

we are given that

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

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

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

using ①, if $\lambda_i = \frac{1}{M}$ and $x_i, \epsilon_i(x)$

let $f(x) = x^2$, $x = \sum_{i=1}^{M} \lambda_i x_i$

In this case, we can have equation ① as

$$\left(\sum_{i=1}^{M} \frac{1}{M} \epsilon_i(x)\right)^2 \leq \sum_{i=1}^{m} \frac{1}{M} \epsilon_i(x)^2$$

as it is true for x, it is true for $E(x)$,

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

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

using this, we can prove that

$\boxed{E_{agg} \leq E_{avg}}$, hence proved.

Question 3:-

We are given that
Hypothesis for boolean Classification problem

$$H(x) = \text{sign}\left(\sum_{t=1}^{N} \alpha_t h_t(x)\right) \rightarrow (1)$$

The weight for point i at time step t+1 is

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times e^{-\alpha_t h_t(i) y(i)} \rightarrow (2)$$

Here, $D_t(i) \rightarrow$ normalized weight for point i at time t

$Z_t \rightarrow$ normalization factor at step t

$y(i) \rightarrow$ true label for point i
$\alpha_t \rightarrow$ Voting parameter for hypothesis h_t

$h_t(i) \rightarrow$ hypothesis function for i at time t

$$D_1 = \frac{1}{N} \text{ for all points } i \rightarrow (3)$$

The error E_t for Adaboost process can be measured with respect to D_t as

$$E_t = \sum_{i: h_t(i) \neq y(i)} D_t(i) \rightarrow (4)$$

as the error at a particular time t is sum of weights corresponding to all points i which are mis classified i.e, $h_t(i) \neq y_i$

Now, from 2,

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

Since $h_t(i)$ and y_i are in both in $\{-1, 1\}$

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

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

where $f_t(i) = -\sum_{j=1}^{t} \alpha_j h_j(i)$

Now, total training error of $h(n)$

$$T_H = \frac{1}{N} \sum_{i = h(i) \neq y(i)} 1,$$

average of mis-classified points also known as accuracy loss

Now, since $H(i) = \text{sign}(f(i))$

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

because, for misclassified points y_i and $f(i)$ would have opposite signs and hence $y_i f(i) \leq 0$

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

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

D_{t+1} is a probability distribution

hence $\sum_i D_{(t+1)}^{\,i} = 1$

$$\Rightarrow T_H \leq \prod_t z_t \rightarrow (6)$$

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

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

because for $h_t(i) = y(i)$

$h_t(i) \times y(i) = 1$ as both would be 1

or both would be -1

If $h_t(i) \neq y(i)$, then $h_t(i) \times y(i) = -1$ as only one of them would be 1.

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

$$Z_t = e^{-\alpha_t}(1 - \varepsilon_t) + e^{\alpha_t} \varepsilon_t \quad \text{(from 4)}$$

Now, to minimize cost T_H,

α_t comes out to be $\frac{1}{2} \frac{1-\varepsilon_t}{\varepsilon_t}$

putting it in above eqn,

$$Z_t = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$$

$\varepsilon_t = \frac{1}{2} - \gamma_t$ gives

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

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

Since, $1+x \leq e^{x} \quad \forall \, x \in \mathbb{R}$

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

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

putting Z_t in 6. gives

$$T_H \leq \prod_t Z_t$$

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

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

Hence proved.