Assignment 3 Part 1

Randy Suarez Rodes rxs179030

Derek Carpenter dxc190017

(1)

Q1

Let $\epsilon_i(X) = f(x) - h_i(x) \quad i=1, \ldots, M$

where h_i is the model created using the i^{th} bootstrap sample

Let $E(\epsilon_i(x)^2) = E[(f(x) - h_i(x))^2]$

and $E_{avg} = \frac{1}{M} \sum_{i=1}^{M} E(\epsilon_i(X)^2)$

Consider $h_{agg}(X) = \frac{1}{M} \sum_{i=1}^{M} h_i(X)$

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

which can be written as

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

For simplicity we will write $\epsilon_i(X)$ as ϵ_i

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

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

linearity
of Expected value

$$= \frac{1}{m^2} E\left[(\epsilon_1 + \epsilon_2 + \dots + \epsilon_m)(\epsilon_1 + \epsilon_2 + \dots + \epsilon_m)\right]$$

$$= \frac{1}{m^2} E\left[\epsilon_1^2 + \sum_{k \neq 1} \epsilon_1 \epsilon_k + \epsilon_2^2 + \sum_{k \neq 2} \epsilon_2 \epsilon_k + \dots + \epsilon_m^2 + \sum_{k \neq m} \epsilon_m \epsilon_k\right]$$

$$= \frac{1}{m^2} E\left[\sum_{i=1}^{m} \epsilon_i^2 + \sum_{k \neq 1} \epsilon_1 \epsilon_k + \dots + \sum_{k \neq m} \epsilon_m \epsilon_k\right]$$

$$= \frac{1}{m^2}\left(\sum_{i=1}^{m} E(\epsilon_i^2) + \sum_{k \neq 1} E(\epsilon_1 \epsilon_k) + \dots + \sum_{k \neq m} E(\epsilon_m \epsilon_k)\right)$$

linearity
of expected value

but $\sum_{k \neq t} E(\epsilon_t \epsilon_k) = 0$

since $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$

So $E_{agg} = \frac{1}{m} \cdot \frac{1}{m} \sum_{i=1}^{m} E(\epsilon_i^2) = \frac{1}{m} E_{avg}$.

Q_2 Let f be a convex function:

then $f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i)$ (Jensen's Inequality)

Consider $\sum \lambda_i x_i$ as $\sum_{i=1}^{m} \frac{1}{m} \epsilon_i(x)$ and $f = x^2$

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

by the Jensen's Inequality

and it holds for all values of x, then it will also hold for the expected value over x

So $\underbrace{E\left(\sum \frac{1}{m} \epsilon_i(x)\right)^2}_{E_{agg}} \leq E\left(\sum_{i=1}^{m} \frac{1}{m} \epsilon_i^2(x)\right)$

$$= \underbrace{\frac{1}{m} \sum}_{\text{linearity of } E} E\left(\epsilon_i^2(x)\right) = E_{avg}$$

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

Q_3 Let $\overset{T}{\text{error}}(H) = \frac{1}{N}\sum_{i=1}^{N} I(H(i) \neq y(i))$

be the overall training error at the end of T steps

we will prove that $\overset{T}{\text{error}}(H) \leq e^{-2\sum_{i=1}^{T} \gamma_t^2}$

where $\epsilon_t = \frac{1}{2} - \gamma_t$

First we have that:

$$\underbrace{D_{t+1}(i) = \frac{D_t(i)\cdot e^{-\alpha_t h_t^{(i)} y(i)}}{Z_t}}_{\text{by definition}} \underbrace{=}_{\text{expanding } D_t(i)} \frac{D_{t-1}^{(i)} e^{-\alpha_{t-1} h_{t-1}^{(i)} y(i)} \cdot e^{-\alpha_t h_t^{(i)} y(i)}}{Z_{t-1} Z_t}$$

$$= \cdots = \frac{D_1 \cdot \prod_{k=1}^{t} e^{-\alpha_k h_k(i) y(i)}}{\prod_{k=1}^{t} Z_k} \underbrace{=}_{D_1 = \frac{1}{N}} \frac{1}{N} \cdot \frac{e^{-y(i)\sum_{k=1}^{t} \alpha_k h_k^{(i)}}}{\prod_{k=1}^{t} Z_k}$$

So $D_{t+1}^{(i)} = \frac{1}{N} \cdot \frac{e^{-y(i)\sum_{k=1}^{t} \alpha_k h_k(i)}}{\prod_{k=1}^{t} Z_k}$

Now:

$$Z_K = \sum_{i=1}^{N} D_K(i) \cdot e^{-\alpha_K h_K(i) y(i)}$$ (by definition

$$= \sum_{i:\, h_K(i) = y(i)} D_K(i) \cdot e^{-\alpha_K} + \sum_{i:\, h_K(i) \neq y(i)} D_K(i)\, e^{\alpha_K}$$

$$= e^{-\alpha_K} \underbrace{\sum_{i:\, h_K(i) = y(i)} D_K(i)}_{1-\epsilon_K} + e^{\alpha_K} \underbrace{\sum_{i:\, h_K(i) \neq y(i)} D_K(i)}_{\epsilon_K}$$

since $\epsilon_K = \sum_{i=1}^{N} D_K(i)\, I(h_K(i) \neq y(i))$

So $Z_K = e^{-\alpha_K}(1-\epsilon_K) + e^{\alpha_K}\epsilon_K$

$$= \frac{\sqrt{\epsilon_K}}{\sqrt{1-\cancel{\epsilon_K}}}(1-\cancel{\epsilon_K}) + \frac{\sqrt{1-\epsilon_K}}{\sqrt{\cancel{\epsilon_K}}} \cdot \cancel{\epsilon_K}$$

$$\alpha_K = \frac{1}{2}\ln\left(\frac{1-\epsilon_K}{\epsilon_K}\right) = \ln\frac{\sqrt{1-\epsilon_K}}{\sqrt{\epsilon_K}}$$

$$= \sqrt{\epsilon_K(1-\epsilon_K)} + \sqrt{(1-\epsilon_K)\epsilon_K}$$

$$= 2\sqrt{\epsilon_K(1-\epsilon_K)}$$

$$= 2\sqrt{\left(\frac{1}{2}-\gamma_K\right)\left(\frac{1}{2}+\gamma_K\right)} = 2\sqrt{\frac{1}{4}-\gamma_K^2}$$

$$\epsilon_K = \frac{1}{2} - \gamma_K$$

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

Finally:

$$\overset{T}{\text{error}}(H) = \frac{1}{N}\sum_{i=1}^{N} I(H(i) \neq Y(i))$$

$$= \frac{1}{N}\sum_{i=1}^{N} I\left[Y(i)\sum_{k=1}^{T}\alpha_k h_k(i)\right]$$

$$\leq \underbrace{\frac{1}{N}}_{D_1(i)}\sum_{i=1}^{N} e^{-Y(i)\cdot\sum_{k=1}^{T}\alpha_k h_k(i)}$$

Since $e^{-x} > 1$ if $x < 0$

$$= \sum_{i=1}^{N} D_1(i)\cdot e^{-Y(i)\sum_{k=1}^{T}\alpha_k h_k(i)}$$

$$= \sum_{i=1}^{N} D_{T+1}(i)\prod_{k=1}^{T} Z_k = \prod_{k=1}^{T} Z_k$$

Since: $D_{T+1}(i) = \frac{1}{N}\frac{e^{-Y(i)\sum\alpha_k h_k(i)}}{\prod_{k=1}^{T} Z_k}$

$$\sum_{i=1}^{N} D_{T+1}(i) = 1$$

$$= \prod_{k=1}^{T}\sqrt{1-4\gamma_k^2} \leq \prod_{k=1}^{T} 1-4\gamma_k^2 \leq \prod_{k=1}^{T} 1-2\gamma_k^2$$

$$\leq \prod_{k=1}^{T} e^{-2\gamma_k^2} = e^{-2\sum_{k=1}^{T}\gamma_k^2}$$

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

So $\overset{T}{\text{error}}(H) \leq e^{-2\sum_{k=1}^{T}\gamma_k^2}$