Machine Learning - HW3

Faizaan Charania - 111463646

Q1 Given: $H(x) = \operatorname{sgn}\left\{\sum_{t=1}^{T} \alpha_t h_t(x)\right\} = \operatorname{sgn}\{f(x)\}$

where,

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$$

To show:

$$\epsilon_{training} = \frac{1}{N}\sum_{j=1}^{N} \delta(H(x^j) \neq y^j) \leq \frac{1}{N}\sum_{j=1}^{N} \exp(-f(x^j)y^j)$$

where $\delta(H(x^j) \neq y^j)$ is 1 if $H(x^j) \neq y^j$ & 0 otherwise

Let's consider different output possibilities.

y^j	$f(x^j)$	$H(x^j)$	$I(H(x^j) \neq y^j)$	$\exp(-f(x^j)y^j)$
1	>0	1	0	≥ 0
-1	>0	1	1	≥ 1
1	<0	-1	1	≥ 1
-1	<0	-1	0	≥ 0

We observe that $\exp(-f(x^j)y^j)$ is ≥ 1 when $f(x^j)y^j \leq 0$ & at least ≥ 0 otherwise.

From the table, we notice that

$$I(H(x^j) \neq y^j) \leq \exp(-f(x^j)y^j)$$

always.

$$\therefore \frac{1}{N}\sum_{j=1}^{N} I(H(x^j) \neq y^j) \leq \frac{1}{N}\sum_{j=1}^{N} \exp(-f(x^j)y^j)$$

2. Given: $$w_j^{t+1} = \frac{w_j^{(t)} \exp(-\alpha_t y^j h_t(x^j))}{Z_t}$$

where $$Z_t = \sum_{j=1}^{N} w_j^t \exp(-\alpha_t y^j h_t(x^j))$$

To prove: $$\frac{1}{N}\sum_{j=1}^{N} \exp(-f(x^j)y^j) = \prod_{t=1}^{T} Z_t$$

We know, $$w_j^{t+1} = \frac{w_j^{(t-1)} \exp(-\alpha_{t-1} y^j h_{t-1}(x^j)) \exp(-\alpha_t y^j h_t(x^j))}{Z_t \cdot Z_{t-1}}$$

$$\therefore w_j^{t+1} = \frac{w_j^1 \exp\left(\sum_{t=1}^{T} -\alpha_t h_t(x^j) y^j\right)}{\prod_{t=1}^{T} Z_t}$$

$$= \frac{w_j^1 \exp(-f(x^j) y^j)}{\prod_t Z_t}$$

$$= \frac{\exp(-f(x^j) \cdot y^j)}{N \prod_t Z_t}$$

$\because w_j^1 = \frac{1}{N}$ [all weights are initialized at $\frac{1}{N}$]

Summing over the data points on both sides

$$\sum_{j=1}^{N} \omega_j^{(t+1)} = \sum_{j=1}^{N} \frac{\exp(-f(x^j)\cdot y^j)}{N \prod_t Z_t}$$

[$\because \omega_j^t$ is a distribution over j]

$$\therefore 1 = \sum_{j=1}^{N} \frac{\exp(-f(x^j)\cdot y^j)}{N \prod_t Z_t}$$

$$\therefore \boxed{\frac{1}{N} \sum_{j=1}^{N} \exp(-f(x^j)\cdot y^j) = \prod_{t=1}^{T} Z_t}$$

3) $$\epsilon_t = \sum_{j=1}^{N} \omega_j^t \, \delta(h_t(x^j) \neq y^j)$$

$$\therefore Z_t = (1-\epsilon_t)\cdot \exp(-\alpha_t) + \epsilon_t \exp(\alpha_t) \quad \text{—(i)}$$

a) Minimize Z_t

Differentiate Z_t & equate to 0

$$\therefore \frac{\partial Z_t}{\partial \alpha_t} = 0$$

$$-\exp(-\alpha_t)(1-\epsilon_t) + \epsilon_t \exp(\alpha_t) = 0$$

$$\therefore [\exp(\alpha_t)]^2 = \frac{1-\epsilon_t}{\epsilon_t}$$

$$\therefore \boxed{\alpha_t^{opt} = \frac{1}{2}\log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)} \quad \leftarrow (ii)$$

put (ii) in (i)

$$\therefore Z_t^{opt} = (1-\epsilon_t)\cdot \exp\left(-\frac{1}{2}\log\left[\frac{1-\epsilon_t}{\epsilon_t}\right]\right) + \epsilon_t \times$$

$$\exp\left[\frac{1}{2}\log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)\right]$$

$$= (1-\epsilon_t)\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{-1/2} + \epsilon_t\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{1/2}$$

$$= \left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{-1/2} - \epsilon_t\left[\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{-1/2} - \left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{1/2}\right]$$

$$= \left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{-1/2} - \epsilon_t\left[\frac{\sqrt{\epsilon_t}}{\sqrt{1-\epsilon_t}} \times \left[\frac{2\epsilon_t - 1}{\epsilon_t}\right]\right]$$

$$= \left(\frac{\epsilon_t}{1-\epsilon_t}\right)^{1/2}(2-2\epsilon_t)$$

$$\therefore \boxed{Z_t^{opt} = 2\sqrt{\epsilon_t(1-\epsilon_t)}}$$

b) $\epsilon_t = \frac{1}{2} - \gamma_t$

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

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

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

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

Taking log on both sides.

$$\log Z_t = \frac{1}{2}\log(1 - 4\gamma_t^2)$$

$$\leq \frac{1}{2}(-4\gamma_t^2)$$

$\left[\because \log(1-x) \leq -x \text{ for } 0 \leq x \leq 1\right]$

$$\therefore \boxed{Z_t \leq \exp(-2\gamma_t^2)}$$

$$\therefore \epsilon_{training} \leq \prod_t Z_t \leq \exp\left(-2\sum_{t=1}^{T}\gamma_t^2\right)$$

— (iii)

c) To prove

$$\epsilon_{training} \leq \exp(-2T\gamma^2)$$

if each classifier is better than random.

$\Rightarrow$ Sol^n :-

Let's assume that all weak classifier are better than random,

i.e. $\epsilon_t \leq 0.5 \quad \forall t$

$\therefore \gamma_t \geq 0.5 \quad [\gamma = 0.5]$

from (iii)

$$\epsilon_{training} \leq \exp\left(-2 \sum_{t=1}^{T} \gamma_t^2\right)$$

$$\leq \exp\left(-2 \sum_{t=1}^{T} \gamma^2\right)$$

$$\boxed{\epsilon_{training} \leq \exp(-2T\gamma^2)}$$