CS 4375

ASSIGNMENT ___3___

Names of students in your group:

Madison Humphries (MRH210006)

Number of free late days used: ____________

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

1. Bagging Models

We are given $E_{agg}(x) = E[\{\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)\}^2]$

Now let's expand the expression:

$$E_{agg}(x) = E[\{\frac{1}{M^2}(\sum_{i=1}^{M} \epsilon_i(x)^2 + 2\sum_{1 \le i < j \le M} \epsilon_i(x)\epsilon_j(x)\}]$$

Take expectation & use linearity of expectation:

$$E_{agg}(x) = \frac{1}{M^2}(\sum_{i=1}^{M} E[\epsilon_i(x)^2] + 2\sum_{1 \le i < j \le M} E[\epsilon_i(x)\epsilon_j(x)])$$

Given the errors are uncorrelated: $E(\epsilon_i(x)\epsilon_j(x)) = 0$ for all $i \neq j$

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

We are given $E_{avg} = \frac{1}{M}\sum_{i=1}^{M} E[\epsilon_i(x)^2]$

Plug in E_{avg} into $E_{agg}(x)$

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

$\boxed{\text{Thus we have proven } E_{agg} = \frac{1}{M} E_{avg}}$

2. Jensen's Inequality

We are given Jensen's Inequality

$$f(\sum_{i=1}^{M} \lambda_i x_i) \le \sum_{i=1}^{M} \lambda_i f(x_i)$$

Let's take $\lambda_i = \frac{1}{M}$ to the real numbers $x_i = \epsilon_i(x)$

$$f(\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)) \le \frac{1}{M}\sum_{i=1}^{M} f(\epsilon_i(x))$$

That reduces to

$$(\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x))^2 \le \frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)^2$$

Take Expectation of both sides & Simplify

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

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

3. Adaboost

For each training point i consider the weighted sum $\sum_{t=1}^{T} a_t h_t(i)$

- If $H(x_i) \neq y(i)$ then $\text{sign}(\sum_{t=1}^{T} a_t h_t(i))$ disagrees with $y(i)$, hence

$$y(i)\sum_{t=1}^{T} a_t h_t(i) \leq 0 \quad \& \quad -y(i)\sum_{t=1}^{T} a_t h_t(i) \geq 0$$

Thus $\exp(-y(i)\sum_{t=1}^{T} a_t h_t(i)) \geq 1$

Since $1\{H(x_i) \neq y(i)\} = 1$ in this case we have

$\star$ $1\{H(x_i) \neq y(i)\} \leq \exp(-y(i)\sum_{t=1}^{T} a_t h_t(i))$

- If $H(x_i) = y(i)$ then $y(i)\sum_{t=1}^{T} a_t h_t(i) > 0$

Thus $\exp(-y(i)\sum_{t=1}^{T} a_t h_t(i)) < 1$

Since $1\{H(x_i) \neq y(i)\} = 0$ in this case we have

$\star$ $1\{H(x_i) \neq y(i)\} \leq \exp(-y(i)\sum_{t=1}^{T} a_t h_t(i))$

Therefore the inequality ($\star$) holds for every i.
Averaging over $i = 1, \ldots, N$ gives:

(1) $\frac{1}{N}\sum_{i=1}^{N} 1\{H(x_i) \neq y(i)\} \leq \frac{1}{N}\sum_{i=1}^{N} \exp(-y(i)\sum_{t=1}^{T} a_t h_t(i))$

Now consider the weight update formula $D_{t+1}(i) = \frac{D_t(i)}{Z_t} e^{-a_t h_t(i) y(i)}$

Unroll the expression from $t=1$ to $t=T$, using $D_1(i) = \frac{1}{N}$

$$D_{T+1}(i) = \frac{1}{N} \frac{e^{-\sum_{t=1}^{T} a_t h_t(i) y(i)}}{\prod_{t=1}^{T} Z_t} = \frac{1}{N} \frac{e^{-y(i)\sum_{t=1}^{T} a_t h_t(i)}}{\prod_{t=1}^{T} Z_t}$$

Sum over i & use $\sum_{t=1}^{N} D_{T+1}(i) = 1$

$$1 = \sum_{i=1}^{N} D_{T+1}(i) = \frac{1}{N} \frac{\sum_{i=1}^{N} e^{-y(i)\sum_{t=1}^{T} a_t h_t(i)}}{\prod_{t=1}^{T} Z_t}$$

Rearrange

$$\frac{1}{N} \sum_{i=1}^{N} \exp(-y(i) \sum_{t=1}^{T} a_t h_t(i)) = \prod_{t=1}^{T} Z_t$$

Using (1) from earlier we get

$$\frac{1}{N} \sum_{i=1}^{N} 1\{H(x_i) \neq y(i)\} \leq \prod_{t=1}^{T} Z_t$$

Now by definition we have $Z_t = \sum_{i=1}^{N} D_t(i) e^{-a_t h_t(i) y(i)}$

Partition the sum by misclassified & correctly classified by h_t under D_t

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

as $\sum_{i: h_t(i) = y(i)} D_t(i) = 1 - \varepsilon_t$ & $\sum_{i: h_t(i) \neq y(i)} D_t(i) = \varepsilon$

Adaboost uses $a_t = \frac{1}{2} \ln \frac{1-\varepsilon_t}{\varepsilon_t}$, which simplies Z_t to

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

Now given $\varepsilon_t = \frac{1}{2} - \gamma_t$,

$$Z_t = 2\sqrt{(\tfrac{1}{2} - \gamma_t)(\tfrac{1}{2} + \gamma_t)} = 2\sqrt{\tfrac{1}{4} - \gamma_t^2} = \sqrt{1 - 4\gamma_t^2}$$

Use Inequality $\ln(1-u) \leq -u$ for $0 \leq u \leq 1$ with $u = 4\gamma_t^2$

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

After exponentiating

$Z_t \le e^{-2\gamma_t^2} \rightarrow \prod_{t=1}^{T} Z_t = \exp(-2\sum_{t=1}^{T} \gamma_t^2)$

Now combining everything

$\frac{1}{N}\sum_{i=1}^{N} 1\{H(x_i) \neq y(i)\} \le \prod_{t=1}^{T} Z_t \le \prod_{t=1}^{T} e^{-2\gamma_t^2} = \exp(-2\sum_{t=1}^{T} \gamma_t^2)$

Thus training error of H:

$$\boxed{\frac{1}{N}\sum_{i=1}^{N} 1\{H(x_i) \neq y(i)\} \le \exp(-2\sum_{t=1}^{T} \gamma_t^2)}$$