

CS 4375

ASSIGNMENT 3

Names of students in your group:

Madison Humphries (MRH100006)

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

► Now let's expand the expression:

$$E_{\text{agg}}(x) = E\left[\left(\frac{1}{M^2} \left(\sum_{i=1}^M e_i(x)^2 + 2 \sum_{1 \leq i < j \leq M} e_i(x) e_j(x)\right)\right)^2\right]$$

► Take expectation & use linearity of expectation:

$$E_{\text{agg}}(x) = \frac{1}{M^2} \left(\sum_{i=1}^M E[e_i(x)^2] + 2 \sum_{1 \leq i < j \leq M} E[e_i(x) e_j(x)] \right)$$

► Given the errors are uncorrelated: $E(e_i(x) e_j(x)) = 0$ for all $i \neq j$

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

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

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

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

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

► 2. Jensen's Inequality

► We are given Jensen's Inequality

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

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

$$f\left(\frac{1}{M} \sum_{i=1}^M e_i(x)\right) \leq \frac{1}{M} \sum_{i=1}^M f(e_i(x))$$

► That reduces to

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

Take Expectation of both sides & Simplify

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

$E_{avg}(x)$

$E_{avg}(x)$

Therefore $E_{avg}(x) \leq E_{avg}(x)$

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

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

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

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

Therefore the inequality (*) holds for every i .

Averaging over $i=1, \dots, N$ gives:

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

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

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

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

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

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

Rearrange

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

Using (1) from earlier we get

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

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

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

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

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

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

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

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

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

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

$$z_t \leq e^{-\alpha r_t^2} \rightarrow \prod_{t=1}^T z_t = \exp(-2 \sum_{t=1}^T r_t^2)$$

Now combining everything

$$\frac{1}{N} \sum_{i=1}^N \mathbb{1}\{H(x_i) \neq y_i\} \leq \prod_{t=1}^T z_t \leq \exp(-2 \sum_{t=1}^T r_t^2) = \exp(-2 \sum_{t=1}^T r_t^2)$$

Thus training error of H :

$$\frac{1}{N} \sum_{i=1}^N \mathbb{1}\{H(x_i) \neq y_i\} \leq \exp(-2 \sum_{t=1}^T r_t^2)$$