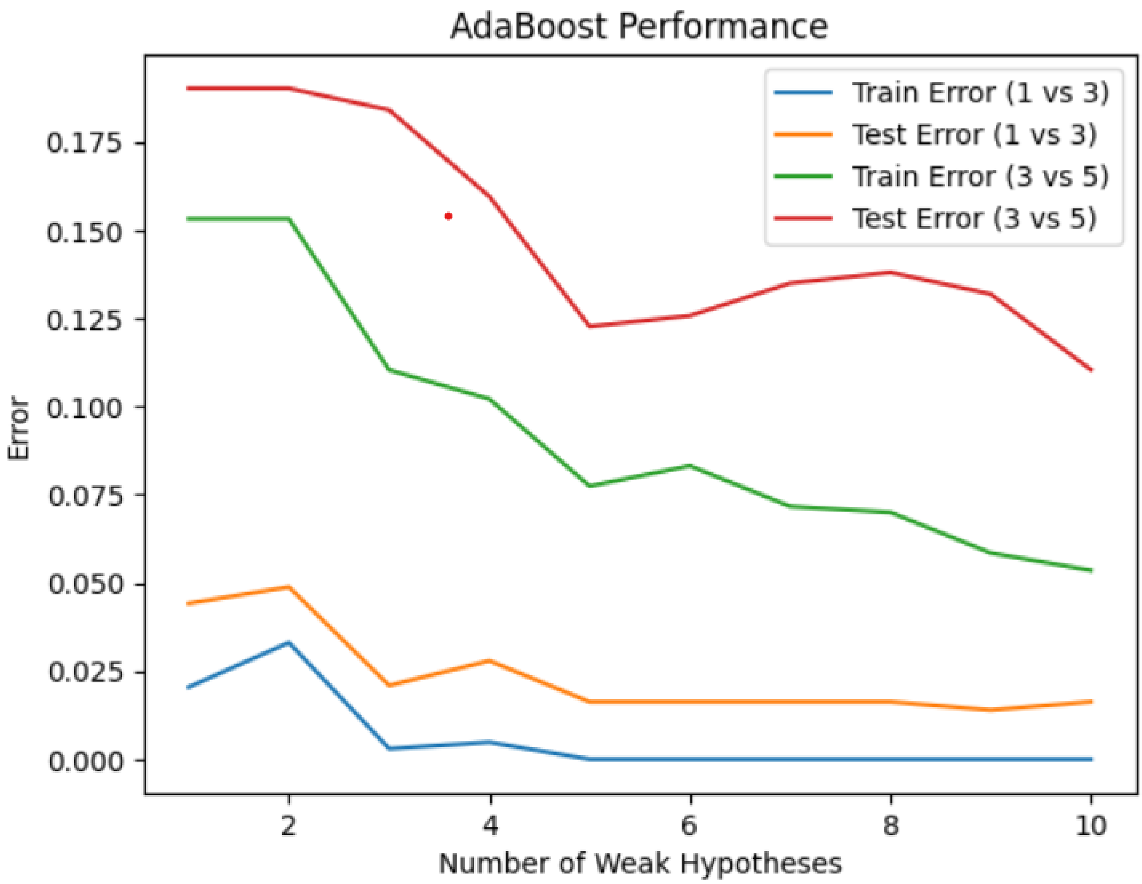
Name: John Stephen Gutam
Email: jgutam@gmu.edu

Problem 1:

Graphical report of the training set error and the test set error as a function of the number of weak hypotheses.



The graph includes four lines, each representing different scenarios.
1. Train error (1 vs 3): The blue line represents the error on the training set when distinguished between classes labeled 1 and 3. As the number of weak hypotheses increases, the error decreases, indicating that the model is learning and improving its predictions on the training data.
2. Test error (1 vs 3): The orange line represents the error on the test set for the same classes 1 and 3. Similar to the training error, the test error also decreases as more weak hypotheses are added, suggesting that the model is generalizing well to unseen data.
3. Train error (3 vs 5): The green line shows the training error for a different pair of classes 3 and 5. The error decreases as more weak hypotheses are included, but it starts to plateau

towards the end, indicating that adding more weak hypotheses is yielding a diminishing return in terms of reducing the training error.
4. Test error (3 vs 5): The red line represents the test error for classes 3 and 5. The error decreases initially as more weak hypotheses are added but then starts to increase slightly after reaching the minimum. This could be an indication of overfitting, where the model performs very well on the training data but less on the test data.

Summary:
The graph demonstrates that AdaBoost can effectively reduce both training and test errors by combining multiple weak hypotheses. However, there is a point beyond which adding more weak hypotheses does not necessarily lead to better performance on the test set. This is an important consideration to avoid overfitting.

For classes 1 and 3, both training and test errors decrease consistently as more weak hypotheses are added, indicating good model performance and generalization. However, for classes 3 and 5, while the training error decreases with more weak hypotheses, the test error starts to increase after reaching a minimum. This suggests that the model may be overfitting to the training data, leading to poorer performance on the test data.

In conclusion, while adding more weak hypotheses can improve the model’s performance, care must be taken to avoid overfitting, which can negatively impact the model’s ability to generalize the unseen data.

Problem 2:

G01413212

Name: John Stephen Gutam

Email: jgutam@gmu.edu

CS 688: Homework 4

2\. In the regression problem where the labels are noisy, that is $y_i = f(x_i) + \epsilon$. Given ϵ is a zero-mean variable with variance σ^2.

Bias - Variance decomposition:

On the hypothesis 'g' of a training set (D). The error is defined as $E_{out}(g^{(D)}) = \mathbb{E}(g^{(D)}(x) - y)^2$ where $\boxed{y = f(x) + \epsilon}$

$$\mathbb{E}_D[L(g^{(D)})] = \mathbb{E}_D[E_{out}(g^{(D)})] = \mathbb{E}_D[\underbrace{E_x(g^D(x) - y)^2}_{\text{Let this be 'z'}}]$$

Lets solve 'z'

$$z = E_D[g^D(x) - y]^2 = E_D[g^D(x) - (f(x) + \epsilon)]^2$$

$$= E_D[g^D(x)^2 + (f(x) + \epsilon)^2 - 2g^D(x) \cdot (f(x) + \epsilon)]$$

$$= E_D[g^D(x)^2 + f(x)^2 + \epsilon^2 + 2f(x)\epsilon - 2 \cdot g^D(x) \cdot f(x) - 2g^D(x)\epsilon]$$

$$= E_D[g^D(x)^2] + f(x)^2 + E(\epsilon^2) + 2f(x)E(\epsilon) - 2Eg^D(x) \cdot f(x) - 2Eg^D(x)E(\epsilon)$$

given $E(\epsilon) = 0$, $E(\epsilon^2) = \sigma^2$

let $\bar{g}(x) = \mathbb{E}[g^D(x)]$

$$= E_D[g^D(x)^2] + f(x)^2 + \sigma^2 + 0 - 2E^D g(x) f(x) - 0$$

$$= E_D[g^D(x)^2] + f(x)^2 + \bar{g}(x)^2 - \bar{g}(x)^2 - 2\bar{g}(x) \cdot f(x) + \sigma^2$$

$$z = \left(E_D[g^{(D)}(x)^2] - \bar{g}(x)^2\right) + \left(\bar{g}(x)^2 - 2\bar{g}(x) f(x) + f(x)^2\right) + \sigma^2$$

$$z = E_D[(g^{(D)}(x) - \bar{g}(x))^2] + (\bar{g}(x) - f(x))^2 + \sigma^2$$

$$\boxed{\text{From: } E[\cancel{x} - \bar{x}]^2 = E[x^2] - \bar{x}^2}$$

Now

$$E_D[L(g^{(D)})] = E_D[E_{out}(g^{(D)})] = E_D[E_x(g^D(x) - f(x))^2]$$

$$= E_x[E_D(g^D(x) - f(x))^2]$$

$$= E_x[z]$$

$$E_D[L(g^{(D)})] = E_x\Big[\underbrace{E_D[(g^{(D)}x - \bar{g}(x))^2]}_{\text{Variance}} + \underbrace{(\bar{g}(x) - f(x))^2}_{\text{bias}} + \sigma^2\Big]$$

So.

$$E_D[L(g^{(D)})] = \text{bias} + \text{Variance} + \sigma^2$$

Hence proved.

Problem 3:

3. The true target function $y=x^2$, the hypothesis space $\mathcal{H}$ has linear functions $h(x)=ax+b$, and each training set we have two examples (x_1, x_1^2), (x_2, x_2^2)

(a) For a linear function $h(x)=ax+b$, the mean hypothesis $\bar{g}(x)$ will have $(\bar{a}, \bar{b})$ and the equation be

$$\boxed{\bar{g}(x) = \bar{a}x + \bar{b}}$$

$$a = \text{slope of the line} = \frac{y_2-y_1}{x_2-x_1} = \frac{x_2^2 - x_1^2}{x_2-x_1} = \frac{(x_2+x_1)\cancel{(x_2-x_1)}}{\cancel{(x_2-x_1)}}$$

$$\boxed{a = (x_1+x_2)} \Rightarrow \boxed{\bar{a} = \mathbb{E}(x_1+x_2)}$$

$$b = \underset{\text{at } (x_1, x_1^2)}{y \text{ intercept}} = y - ax = x_1^2 - ax_1 = x_1^2 - (x_1+x_2)x_1$$
$$= \cancel{x_1^2} - \cancel{x_1^2} - x_1x_2$$

$$\boxed{b = -x_1x_2} \Rightarrow \boxed{\bar{b} = \mathbb{E}(-x_1x_2)}$$

Therefore $\bar{g}(x) = \bar{a}x + \bar{b} = (\mathbb{E}(x_1+x_2))x - \mathbb{E}(x_1x_2)$.

(b) In uniform distribution

pdf (probability density functn) $= P_r[a \le X \le b] = \int_a^b f(x)dx$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for } x<a \ \& \ x>b \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{b-a} = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$ (From wikipedia)

$$E(x^2) = \int_a^b x^2 \frac{dx}{b-a} = \frac{b^3-a^3}{3(b-a)}.$$

$$\bar{a} = \mathbb{E}(x_1+x_2) = \mathbb{E}(x_1) + \mathbb{E}(x_2)$$

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

$$= 0+0$$

$$= 0.$$

as x_1 and x_2 are independent

$\therefore [-1, 1]$ with $a = -1$, $b = 1$

$\therefore E[x] = \frac{b+a}{2}$

$$\bar{b} = \mathbb{E}(-x_1x_2) = -E(x_1x_2) = -\mathbb{E}(x_1)\,\mathbb{E}(x_2)$$

$$= \cancel{-\left[\frac{(1)^3-(-1)^3}{3(1-(-1))}\right]\left[\frac{(1)^3-(-1)^3}{3(1-(-1))}\right]}$$

$E[x] = \frac{b+a}{2}$

$$= \left(\frac{-1+1}{2}\right)\cdot\left(\frac{-1+1}{2}\right)$$

$\cancel{E[x^2] = \frac{b^3-a^3}{3(b-a)}}$

$$= 0.$$

$$\bar{a^2} = E\left[(x_1+x_2)^2\right] = E[x_1^2] + E[x_2^2] + 2E[x_1x_2]$$

$$= \frac{(1)^3-(-1)^3}{3(1-(-1))} + \frac{(1)^3-(-1)^3}{3(1-(-1))} + 2(0)$$

$$= \frac{2}{3(2)} + \frac{2}{3(2)} + 0$$

From $E[x^2] = \frac{b^3-a^3}{3(b-a)}$

$$\boxed{\bar{a^2} = \frac{2}{3}}$$

$$\overline{b^2} = \mathbb{E}\left[(-x_1 x_2)^2\right] = E\left[x_1^2 x_1^2\right] = E[x_1^2]\,E[x_2^2]$$

$$= \frac{2}{3(2)} \cdot \frac{2}{3(2)}$$

$$= \frac{1}{9}$$

$$\boxed{\overline{b^2} = \frac{1}{9}}$$

(c) bias = Squared difference between mean hypothesis and the true function.

$$= \mathbb{E}\left[\left(\bar{g}(x) - f(x)\right)^2\right]$$

$$= \mathbb{E}\left[\left(\bar{a}x + \bar{b} - x^2\right)^2\right]$$

$$= \mathbb{E}\left[\left(0 + 0 - x^2\right)^2\right]$$

$$= \mathbb{E}\left[x^4\right]$$

$$= \int_{-1}^{1} \frac{1}{(b-a)} (x^4)\, dx = \frac{1}{2}\left[x^5/5\right]_{-1}^{1}$$

$$= \frac{1}{10}\left[(1)^5 - (-1)^5\right] = \frac{1}{10}\left[1 - (-1)\right] = \frac{2}{10}$$

$$\boxed{\text{bias} = 1/5} = 0.2$$

$$\text{Variance} = E\left[\left(g(x) - \bar{g}(x)\right)^2\right]$$

$$= E\left[\left(ax+b - (\bar{a}x+\bar{b})\right)^2\right]$$

$$= E\left[\left(ax+b - 0+0\right)^2\right]$$

$$= E\left[(ax+b)^2\right]$$

$$= E\left[a^2x^2 + b^2 + 2axb\right]$$

$$= a^2E[x^2] + E[b^2] + 2ab\,E[x]$$

$$= a^2\left(\frac{(1)^3 - (-1)^3}{3(1-(-1))}\right) + b^2 + 2ab\left(\frac{1+(-1)}{2}\right)$$

$$= a^2\left(\frac{1}{3}\right) + b^2$$

$$= (-1)^2\left(\frac{1}{3}\right) + (1)^2$$

$$\boxed{\text{Variance} = 4/3} = 1.33$$

$$E(x) = \frac{b+a}{2}$$

$$E(x^2) = \frac{b^3 - a^3}{3(b-a)}$$

$[-1, 1]$
a b

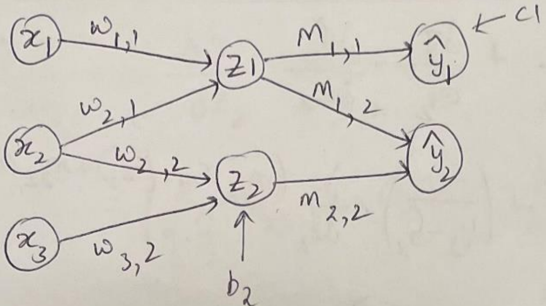
Mean squared Error (E) = bias + variance

$$= \frac{1}{5} + \frac{4}{3}$$

$$= \frac{3+20}{15} = \frac{23}{15} = 1.53$$

Problem 4:

(4)



Given

$$z_1 = \mathrm{ReLU}(w_{1,1} x_1 + w_{2,1} x_2)$$

$$z_2 = w_{2,2} x_2^2 + w_{3,2} x_3 + b_2$$

$$\hat{y}_1 = \sigma(m_{1,1} z_1^3 + c_1)$$

$$\hat{y}_2 = m_{1,2} \sin(z_1) + m_{2,2} \cos(z_2)$$

$$J(\theta) = 2 \log(y_1 - \hat{y}_1) + 2 \log(y_2 - \hat{y}_2)$$

(a) $$\frac{\partial J}{\partial \hat{y}_1} = 2 \cdot \frac{1}{y_1 - \hat{y}_1} (-1) = \frac{-2}{y_1 - \hat{y}_1} \Rightarrow \boxed{\frac{\partial J}{\partial \hat{y}_1} = \frac{-2}{y_1 - \hat{y}_1}}$$

similarly $$\boxed{\frac{\partial J}{\partial \hat{y}_2} = \frac{-2}{y_2 - \hat{y}_2}}$$

Note:
$$\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))$$

(b) (i) $$\frac{\partial J}{\partial z_1} = \frac{\partial J}{\partial \hat{y}_1} \times \frac{\partial \hat{y}_1}{\partial z_1} + \frac{\partial J}{\partial \hat{y}_2} \times \frac{\partial \hat{y}_2}{\partial z_1}$$

$$= \left(\frac{-2}{y_1 - \hat{y}_1}\right) \frac{\partial(\sigma(m_{11} z_1^3 + c_1))}{\partial z_1} + \left(\frac{-2}{y_2 - \hat{y}_2}\right) \times \frac{d(m_{12} \sin(z_1) + m_{22} \cos(z_2))}{\partial z_1}$$

$$= -2 \left[\frac{\hat{y}_1 (1 - \hat{y}_1)(m_{11} 3 z_1^2)}{y_1 - \hat{y}_1} + \frac{m_{12} \cos(z_1)(1)}{y_2 - \hat{y}_2} \right]$$

$$\frac{\partial J}{\partial z_1} = -2 \left[\frac{3 m_{11} z_1^2 \hat{y}_1 (1 - \hat{y}_1)}{y_1 - \hat{y}_1} + \frac{m_{12} \cos(z_1)}{y_2 - \hat{y}_2} \right]$$

(ii) $$\frac{\partial J}{\partial z_2} = \frac{\partial J}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial z_2} + \frac{\partial J}{\partial \hat{y}_2} \cdot \frac{\partial \hat{y}_2}{\partial z_2}$$

$$= \left(\frac{-2}{y_1 - \hat{y}_1}\right)(0) + \left(\frac{-2}{y_2 - \hat{y}_2}\right) \cdot \frac{\partial}{\partial z_2}\left(m_{12} \sin(z_1) + m_{22} \cos(z_2)\right)$$

$$= \frac{-2}{y_2 - \hat{y}_2} m_{22} \left(-\sin(z_2)\right)$$

$$\frac{\partial J}{\partial z_2} = \frac{2}{y_2 - \hat{y}_2} m_{22} \left(\sin(z_2)\right)$$

(iii) $$\frac{\partial J}{\partial m_{1,1}} = \frac{\partial J}{\partial \hat{y}_1} \times \frac{\partial \hat{y}_1}{\partial m_{1,1}} = \frac{\partial J}{\partial \hat{y}_1} \cdot \frac{\partial \left(\sigma(m_{1,1} z_1^3 + c_1)\right)}{\partial m_{1,1}}$$

$$= \left(\frac{-2}{y_1 - \hat{y}_1}\right) \cdot \hat{y}_1 (1 - \hat{y}_1) \cdot z_1^3$$

$$\boxed{\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))}$$
$$= \hat{y}_1 (1 - \hat{y}_1)$$

(iv) $$\frac{\partial J}{\partial m_{1,2}} = \frac{\partial J}{\partial \hat{y}_2} \cdot \frac{\partial \hat{y}_2}{\partial m_{1,2}} = \frac{\partial J}{\partial \hat{y}_2} \cdot \frac{\partial \left(m_{12} \sin(z_1) + m_{22} \cos(z_2)\right)}{\partial m_{1,2}}$$

$$= \left(\frac{-2}{y_2 - \hat{y}_2}\right) \cdot \sin(z_1)$$

(v) $$\frac{\partial J}{\partial c_1} = \frac{\partial J}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial c_1} = \frac{-2}{(y_1 - \hat{y}_1)} \cdot \frac{\partial \left(\sigma(m_{1,1} z_1^3 + c_1)\right)}{\partial c_1}$$

$$= \frac{-2}{(y_1 - \hat{y}_1)} \hat{y}_1 (1 - \hat{y}_1)(1) = \frac{-2 y_1 (1 - \hat{y}_1)}{(y_1 - \hat{y}_1)}$$

(c)

(i) $$\frac{\partial J}{\partial w_{2,2}} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{2,2}}$$

$$= \left(\frac{2}{y_2 - \hat{y}_2} m_{2,2} (\sin(z_2))\right) \cdot \frac{\partial}{\partial w_{2,2}} (w_{2,2} x_2^2 + w_{3,2} x_3 + b_2)$$

$$= \frac{2}{y_2 - \hat{y}_2} m_{2,2} (\sin(z_2)) \cdot x_2^2$$

$$\frac{\partial J}{\partial w_{2,2}} = \frac{2 x_2^2 m_{2,2} \sin(z_2)}{y_2 - \hat{y}_2}$$

(ii) $$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial b_2} = \left[\frac{2}{y_2 - \hat{y}_1} m_{22} (\sin(z_2))\right] \frac{\partial (w_{22} x_2^2 + w_{3,2} x_3 + b_2)}{\partial b_2}$$

$$= \frac{2}{y_2 - \hat{y}_1} m_{22} (\sin(z_2)) \times (1)$$

$$\frac{\partial J}{\partial b_2} = \frac{2}{y_2 - \hat{y}_1} m_{2,2} (\sin(z_2))$$

(iii) $$\frac{\partial J}{\partial w_{1,1}} = \frac{\partial J}{\partial z_1} \times \frac{\partial z_1}{\partial w_{1,1}} = \frac{\partial J}{\partial z_1} \times \frac{\partial \left(\mathrm{ReLU}(w_{11} x_1 + w_{21} x_2)\right)}{\partial w_{1,1}}$$

$$\boxed{\text{where } \frac{\partial}{\partial x} \mathrm{ReLU}(x) = \mathrm{ReLU}'(x) = \begin{cases} 0, & x < 0 \\ 1, & \text{others.} \end{cases}}$$

$$= \frac{\partial J}{\partial z_1} \times \mathrm{ReLU}'(w_{11} x_1 + w_{21} x_2) \cdot x_1$$

$$\frac{\partial J}{\partial w_{1,1}} = \frac{\partial J}{\partial z_1} \times ReLU'(w_{11}x_1 + w_2 x_2) \times x_1$$

$$\boxed{\text{if } x < 0,\ ReLU'(x) = 0}$$

at $x < 0$, $\frac{\partial J}{\partial w_{1,1}} = \frac{\partial J}{\partial z_1} \cdot (0) \cdot x_1 = 0 \Rightarrow \boxed{\frac{\partial J}{\partial w_{11}} = 0 \text{ at } x<0}$

at ~~others~~ others, $\frac{\partial J}{\partial w_{1,1}} = \frac{\partial J}{\partial z_1} (1) \cdot x_1$ if x is others $ReLU'(x) = 1$

$$= \frac{\partial J}{\partial z_1} \cdot (x_1)$$

$$\boxed{\frac{\partial J}{\partial z_1} = -2\left[\frac{3m_{11} z_1^2 \hat{y}_1(1-\hat{y}_1)}{y_1 - \hat{y}_1} + \frac{m_{12}\cos(z_1)}{y_2 - \hat{y}_2}\right](x_1) \quad \text{at } x \geq 0}$$