Ian Borwick – 250950449

iborwick@uwo.ca

CS 4442 – Assignment 1

Contents

Question 1

1a)

Date

1a) $f(w) = w^T x$, where $x \in \mathbb{R}^n$ is an nth-dimensional array

find $\nabla f(w)$ using gradient def:

$$f(w) = w^T x$$

$$= (w_1, w_2, \dots, w_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

def: $\nabla f(w) = \left(\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \dots, \frac{\partial f}{\partial w_n} \right)$

$$\therefore \nabla f(w) = (x_1, x_2, \dots, x_n)$$

$$= x^T$$

$$\boxed{\therefore \nabla f(w) = x^T}$$

1b)

Date

1b) $$f(w) = tr(ww^T A)$$
$$= tr(w^T A w)$$
$$= tr\left(w^T \begin{bmatrix} \sum_{i=1}^{n} a_{1i} w_i \\ \sum_{i=1}^{n} a_{2i} w_i \\ \vdots \\ \sum_{i=1}^{n} a_{ni} w_i \end{bmatrix}\right)$$
$$= \sum_{j=1}^{n} \sum_{i=1}^{n} w_j a_{ji} w_i$$
$$= \sum_{j=1}^{n} \left(a_{jj} w_j^2 + \sum_{i \neq j} w_j a_{ji} w_i\right)$$
$$\frac{\partial}{\partial w_k} = \left[\sum_{j=1}^{n} \left(a_{jj} w_j^2 + \sum_{i \neq j} w_j a_{ji} w_i\right)\right]$$
$$= \sum_{i=1}^{n} w_i a_{i12} + \sum_{i=1}^{n} a_{12i} w_i$$
$$\nabla f(w) = \begin{bmatrix} \frac{\partial}{\partial w_1} \\ \frac{\partial}{\partial w_2} \\ \vdots \\ \frac{\partial}{\partial w_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} w_i a_{i1} \\ \sum_{i=1}^{n} w_i a_{i2} \\ \vdots \\ \sum_{i=1}^{n} w_i a_{in} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{n} a_{1i} w_i \\ \sum_{i=1}^{n} a_{2i} w_i \\ \vdots \\ \sum_{i=1}^{n} a_{ni} w_i \end{bmatrix}$$
$$= A^T w + A w$$
$$= w(A + A^T)$$
$$\boxed{\nabla f(w) = 2AW}$$

Page

1c)

Date

1c) $f(w) = tr(ww^TA)$

find the Hessian matrix H of f

$f(w) = tr(ww^TA)$

$$\frac{\partial f}{\partial w_k} = \sum_{i=1}^{n} w_i a_{ik} + \sum_{i=1}^{n} a_{ki} w_i$$

$$\frac{\partial^2 f}{\partial w_k \partial w_{k'}} = a_{k'k} + a_{kk'}$$

$$\therefore \nabla^2 f(w) = \begin{bmatrix} \frac{\partial}{\partial w_1 \partial w_1} f(w), & \frac{\partial}{\partial w_1 \partial w_2} f(w), & \dots, & \frac{\partial}{\partial w_1 \partial w_n} f(w) \\ \frac{\partial}{\partial w_2 \partial w_1} f(w), & \dots & \dots, & \frac{\partial}{\partial w_2 \partial w_n} f(w) \\ \vdots & & \ddots & \\ \frac{\partial}{\partial w_n \partial w_1} f(w), & \dots & \dots, & \frac{\partial}{\partial w_n \partial w_n} f(w) \end{bmatrix}$$

$$\therefore H = \begin{bmatrix} a_{11}+a_{11}, & a_{21}+a_{12}, & \dots, & a_{n1}+a_{1n} \\ a_{12}+a_{21}, & a_{22}+a_{22}, & \dots, & a_{n2}+a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{1n}+a_{n1}, & \dots & \dots, & a_{nn}+a_{nn} \end{bmatrix}$$

$$\therefore \nabla^2 f(w) = A + A^T$$

Page

1d)

Date

1d) $f(w) = \log(\sigma(w^T x))$

let: $w^T x = a$

$\therefore f(w) = \log(\sigma(a))$

apply chain rule ($f(g(x)) = f'(g(x))\, g'(x)$)

$\frac{d}{dx}\sigma(a) = \sigma(a)(1-\sigma(a))$ $\quad \frac{d}{dx}\log(\sigma(a)) = \frac{1}{\sigma(a)\ln(10)}$

$$\therefore \frac{d}{dx} f(w) = \frac{1}{\sigma(a)\ln(10)} \cdot \sigma(a)(1-\sigma(a))$$

$$= \frac{\sigma(a) - \sigma^2(a)}{\sigma(a)\ln(10)}$$

$$= \frac{1-\sigma(a)}{\ln(10)}$$

$\nabla a = \nabla(w^T x) = w^T$ (from 1a)

$$\nabla f(w) = \frac{1-\sigma(w^T)}{\ln(10)}$$

$\sigma(a) = \frac{1}{1+e^{-a}}$

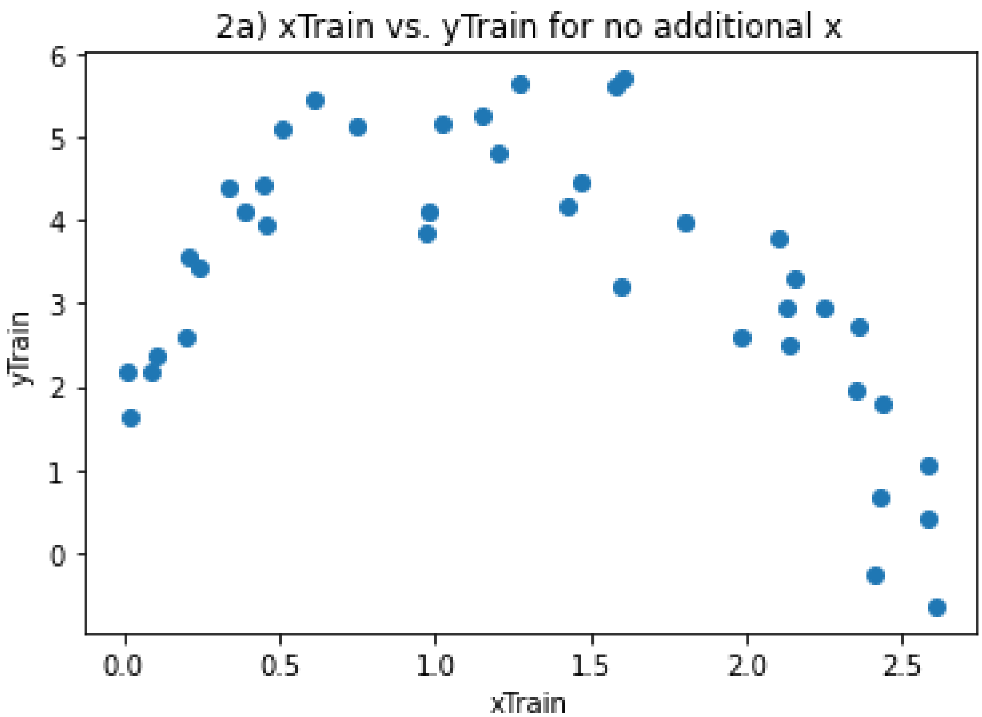
$$= \frac{1 - \frac{1}{1+e^{w^T}}}{\ln(10)}$$

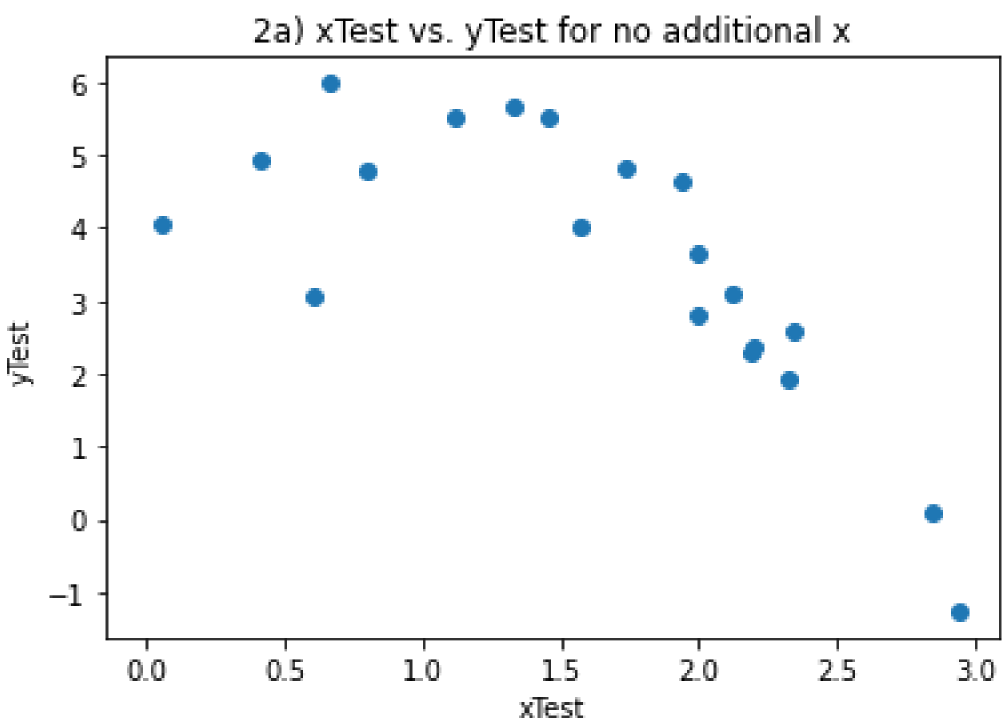
$$\boxed{\nabla f(w) = \frac{e^{w^T}}{(1+e^{w^T})\ln(10)}}$$

Page

Question 2

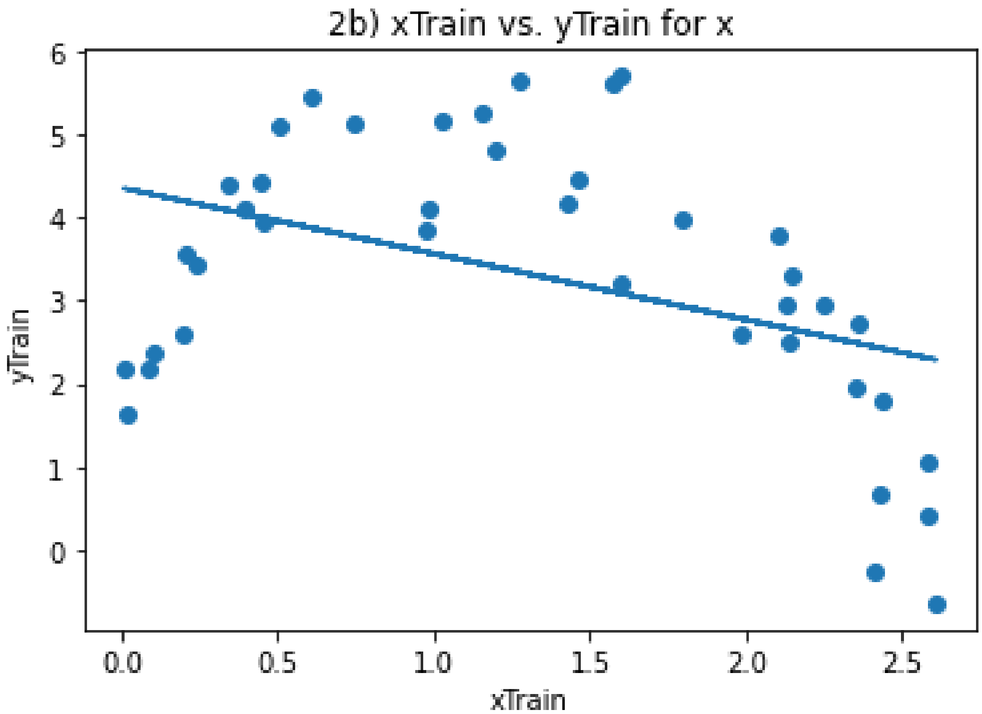
2a)





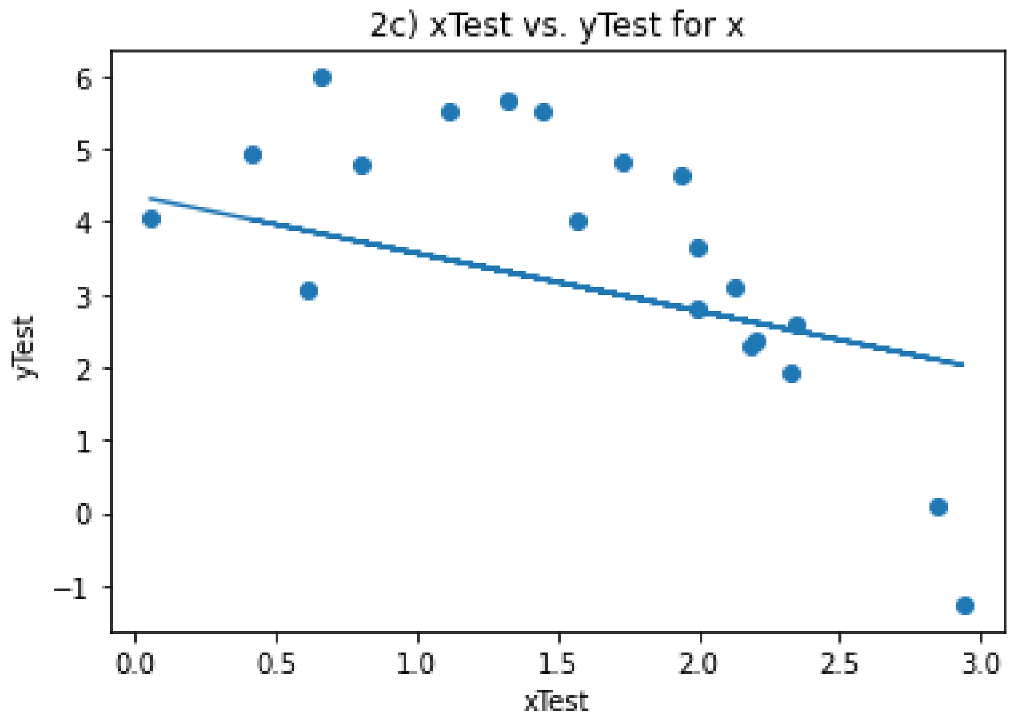
2b)

The training error for this regression model is: [2.17394558]



2c)

The test error for this regression model is: [2.31187535]

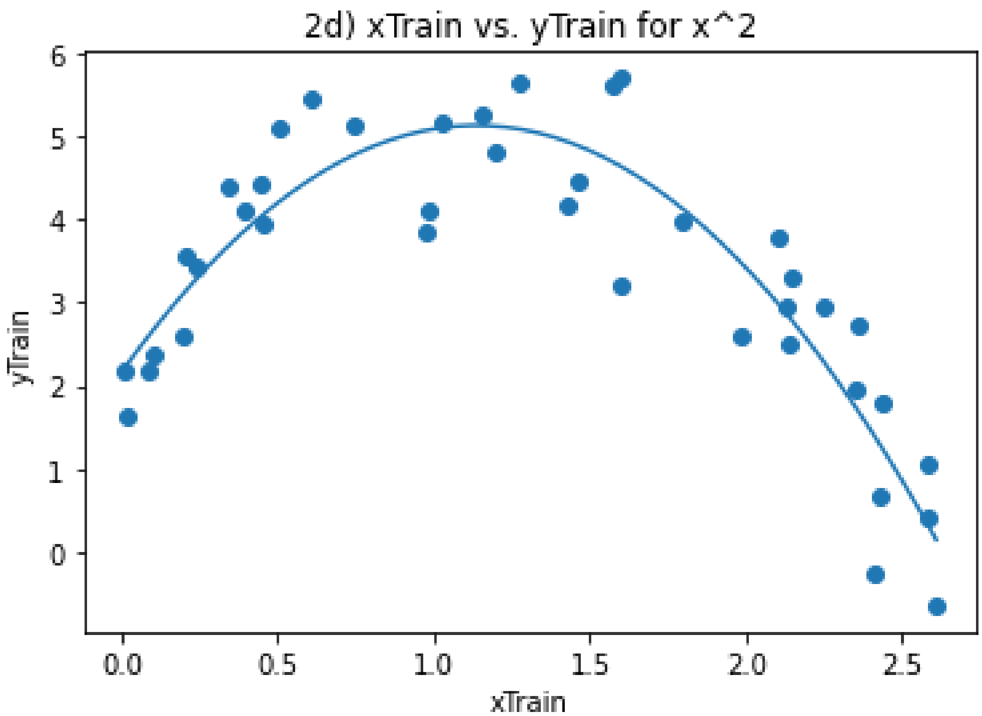


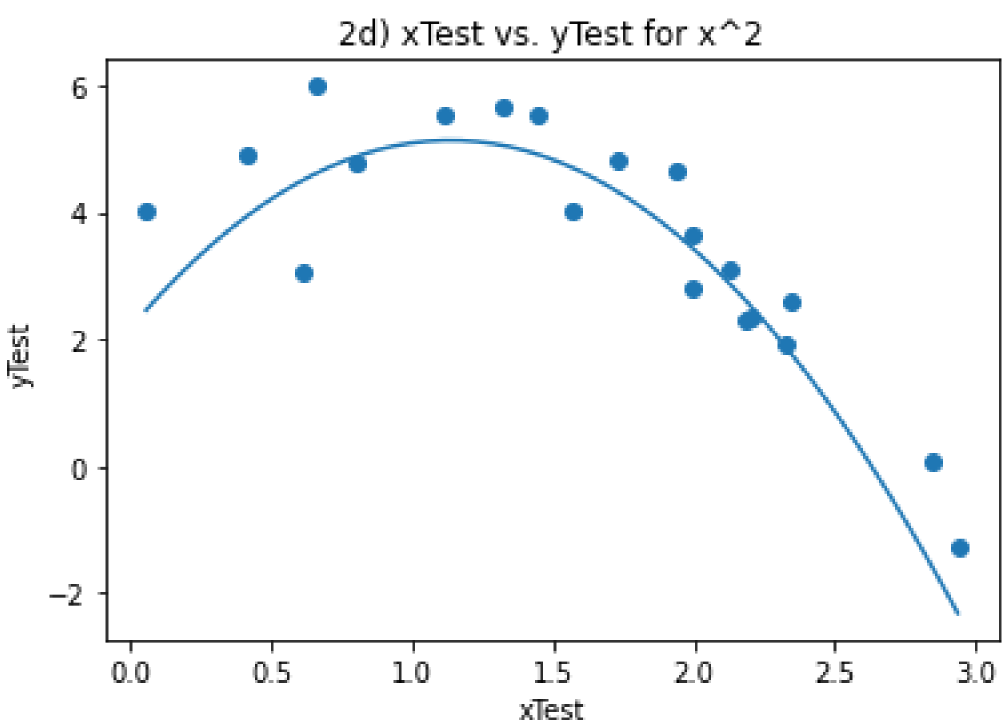
2d)

The training error for the x^2 regression model is: [0.4846845]

The test error for the x^2 regression model is: [0.75736357]

The x^2 regression is a better fit than linear



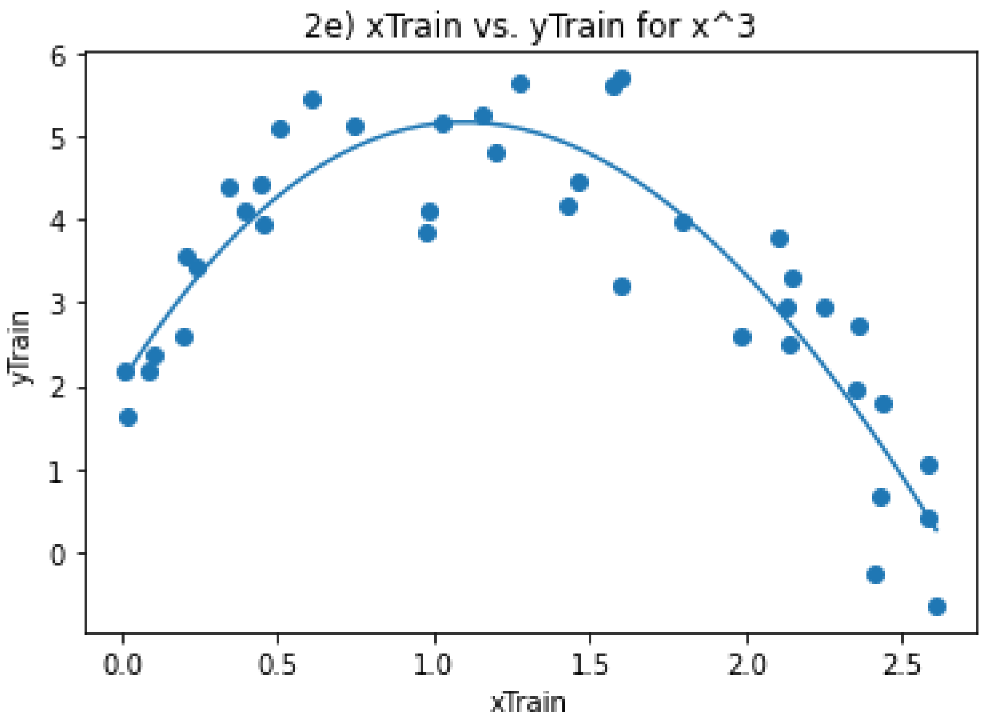


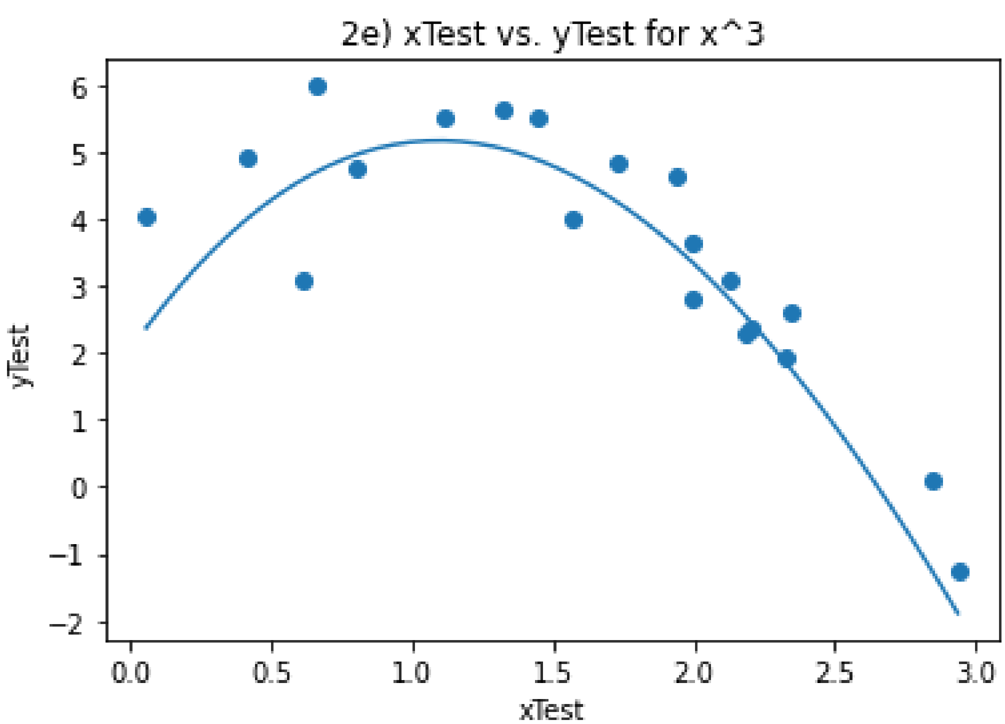
2e)

The training error for the x^3 regression model is: [0.48055213]

The test error for the x^3 regression model is: [0.69112454]

The x^3 regression is a better fit than linear and better than x^2



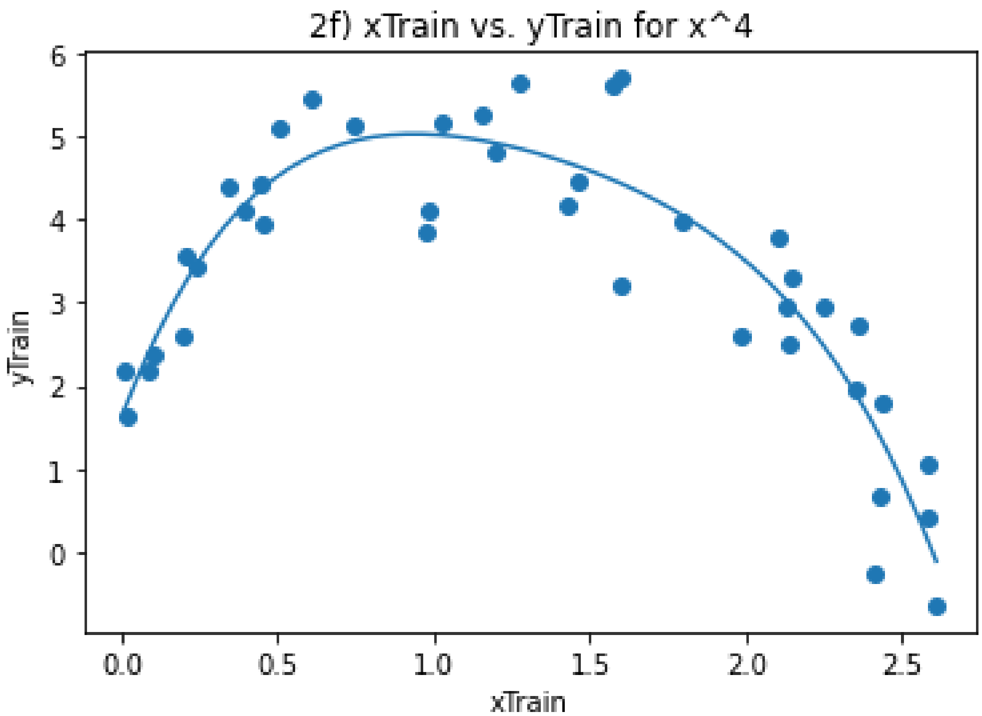


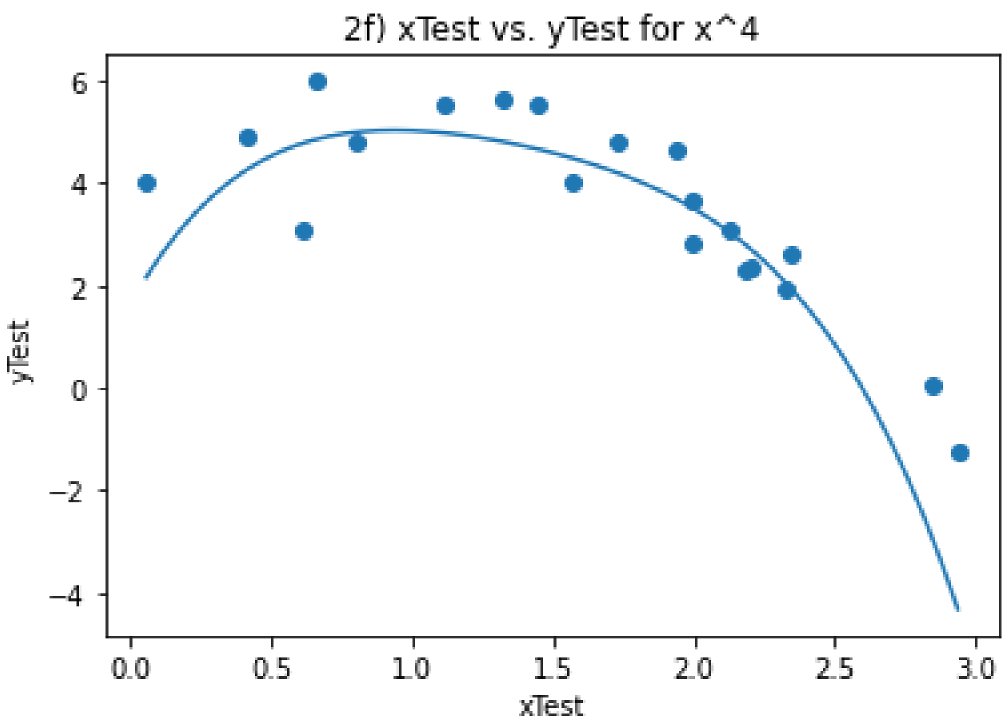
2f)

The training error for the x^4 regression model is: [0.43664763]

The test error for the x^4 regression model is: [1.55846948]

The x^4 regression is a better fit than linear but worse than the x^2 and x^3





Question 3

3a)

Train Error values:

Blue - 0.01 : [0.44766944]

Orange - 0.1 : [0.50764684]

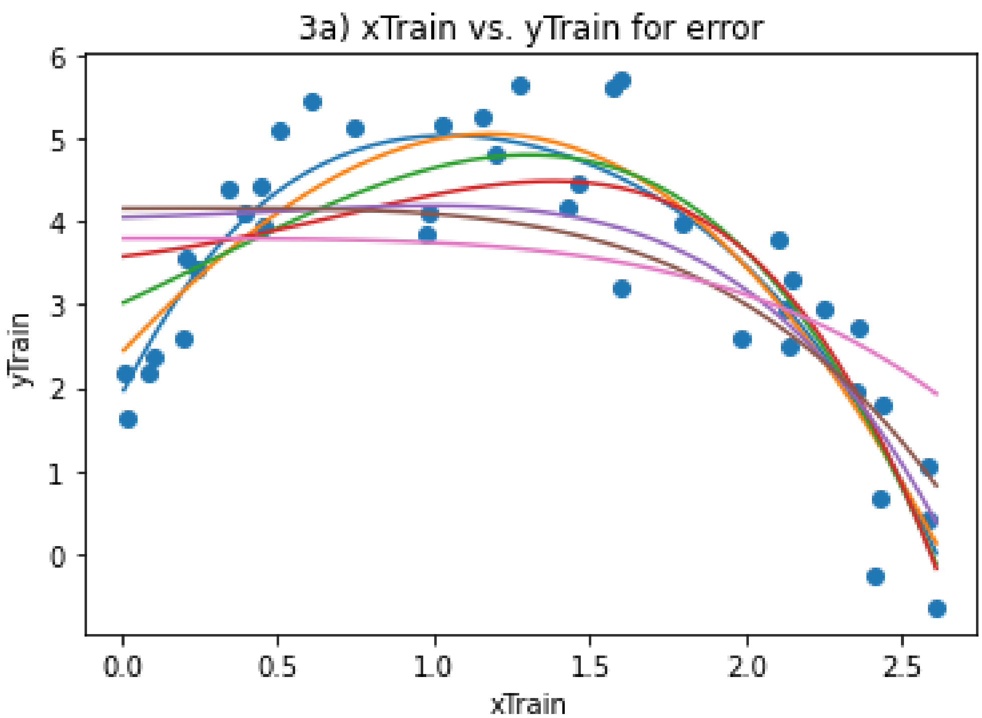
Green- 1 : [0.61438502]

Red - 10 : [0.79919963]

Purple - 100 : [1.06408674]

Brown - 1000 : [1.24504551]

Pink - 10000 : [1.58866293]



Test Error values:

Blue- 0.01 : [1.1142357]

Orange - 0.1 : [0.75921838]

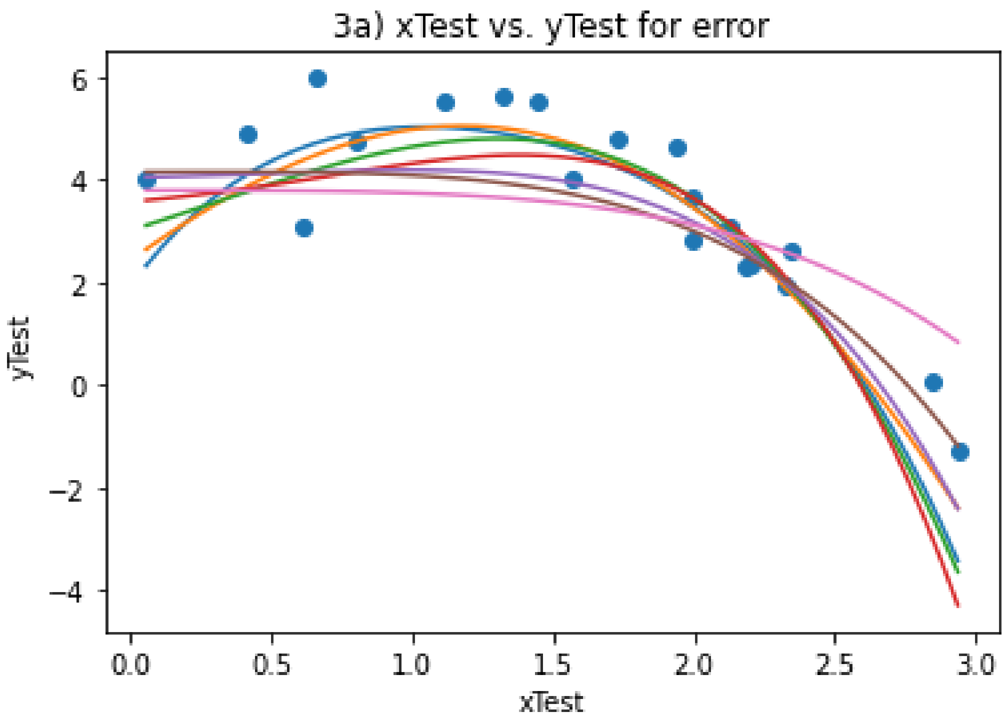
Green - 1 : [1.19209465]

Red - 10 : [1.56971215]

Purple -100 : [0.97921956]

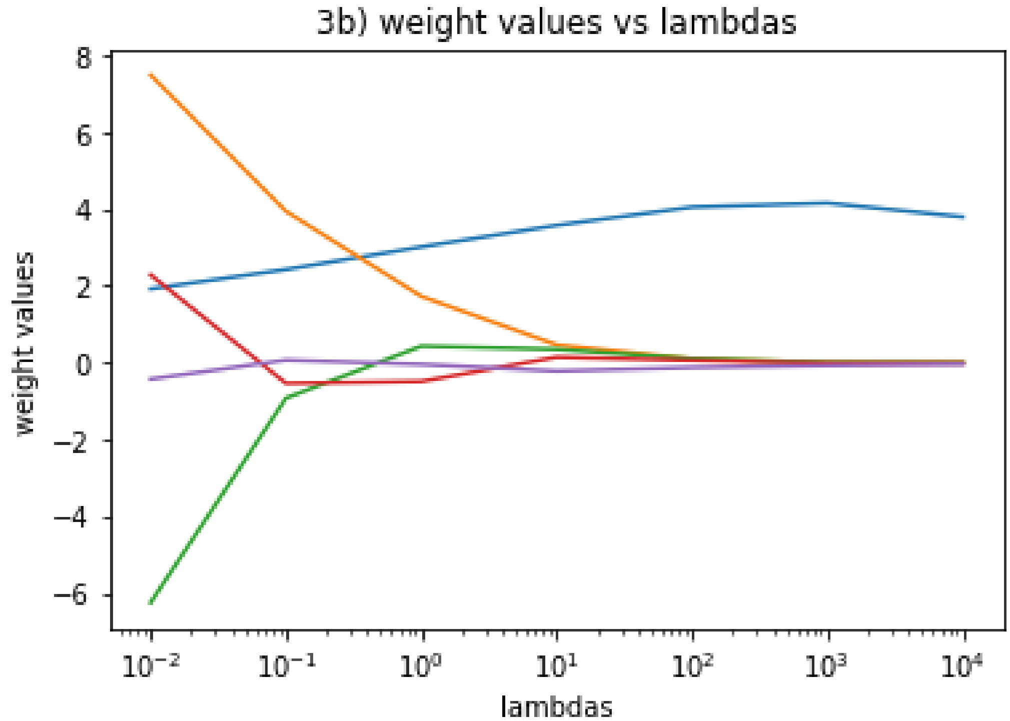
Brown - 1000 : [0.95177873]

Pink - 10000 : [1.50004816]



The best lambda fit is for 0.1 as it had the lowest error for the testing data

3b)


3b) weight values vs lambdas
weight values
lambdas
8
6
4
2
0
−2
−4
−6
10^−2
10^−1
10^0
10^1
10^2
10^3
10^4

3c)

Average Errors

0.01 : [0.56797556]

0.1 : [0.61637959]

1 : [0.74211609]

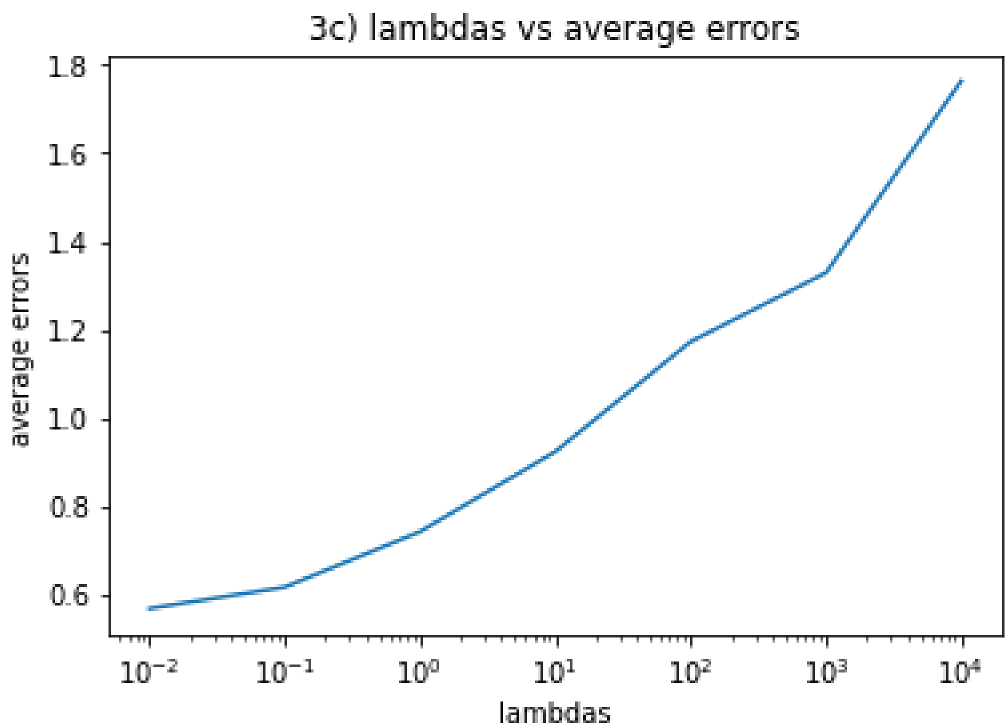
10 : [0.92384521]

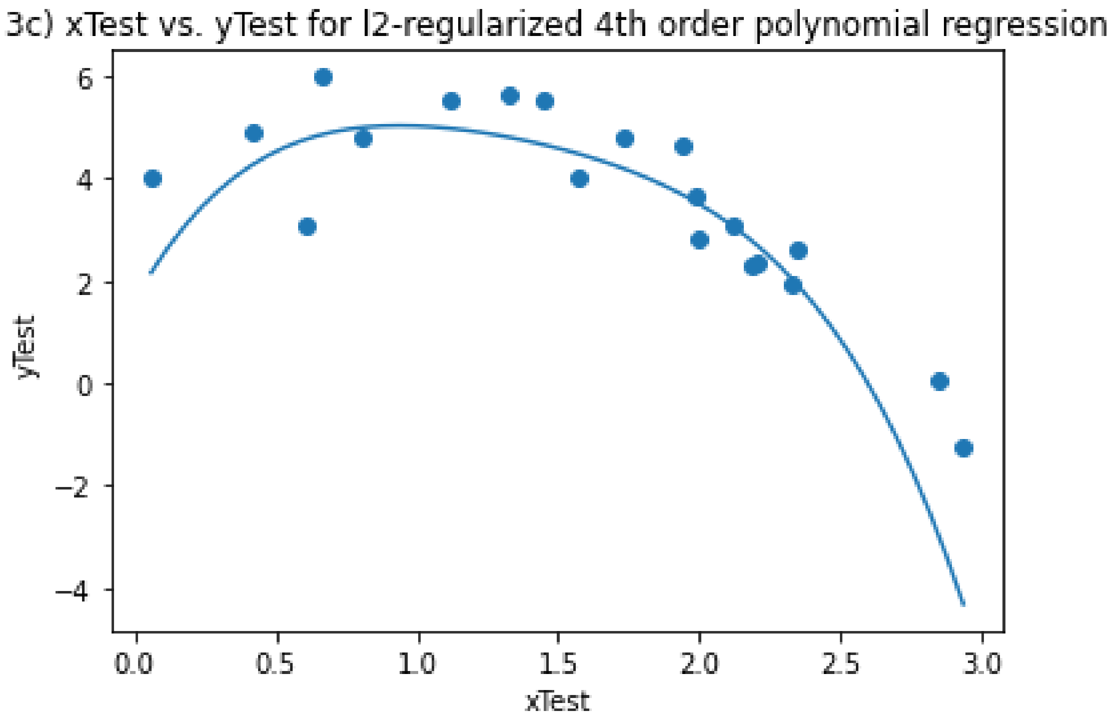
100 : [1.17268211]

1000 : [1.32873231]

10000 : [1.76322053]

The best lambda for c) was 0.01 and for a) it was 0.1 so it did changed




3c) xTest vs. yTest for l2-regularized 4th order polynomial regression
yTest
xTest
6
4
2
0
−2
−4
0.0
0.5
1.0
1.5
2.0
2.5
3.0