CS 4442B Assignment 1

Ositadinma Arimah

250 981 235

Department of Computer Science
Western University

2022-02-16

Ositadinma Arimah
CS 4442B Assignment 1

I.

a.

1a let $f(w) = w^T x b$ where $x \in R^{n \times p}$ is a $n \times p$ matrix and b is a p-dimensional column vector

let $\nabla f(w)$ using the defn of gradient

$$
\begin{aligned}
f(w) &= w^T x b \\
&= (w^T x b)^T \\
&= b^T x^T w
\end{aligned}
$$

~~Then~~, Now $Df(w) = b^T x^T$

$$
\begin{aligned}
\therefore \quad \nabla f(w) &= [D(fw)] \\
&= (b^T x^T)^T \\
&= ((x b)^T)^T \\
&= x b
\end{aligned}
$$

Therefore, $\nabla f(w) = x b$

Ositadinma Arimah
CS 4442B Assignment 1

b.

1b $f(w) = tr(ww^T A)$

using $tr(AB) = tr(BA \dots$

$$f(w) = tr(w^T A w)$$

$$= tr\left([w_1, w_2, \dots w_n] \begin{bmatrix} \sum_{i=1}^{n} a_{1j} w_j \\ \sum_{i=1}^{n} a_{2j} w_j \\ \vdots \\ \sum_{i=1}^{n} a_{nj} w_j \end{bmatrix} \right)$$

$$= \sum_{i=1}^{a} a_{1j} w_1 w_j + \sum_{i=1}^{n} a_{2j} w_2 w_j + \dots + \sum_{i=1}^{a} a_{nj} w_n w_j$$

$$f(w) = a_{11} w_1^2 + a_{12} w_1 w_2 + \dots + a_{1n} w_j w_n + \dots$$

$$d_{21} w_1 w_2 + d_{22} w_2^2 + \dots + a_{2n} w_n w_n + \dots$$

$$+ a_{n1} w_1 w_n + a_{n2} w_2 w_n + \dots + d_{nn} w_n^2$$

$$\nabla f(w) = \left[\frac{\partial f}{\partial w_1}, \frac{df}{w_2}, \frac{df}{w_3}, \dots, \frac{df}{w_n} \right]$$

$$= (2a_{11} w_1 + a_{12} w_1 + \dots + a_{1n} w_n + a_{21} w_{21}$$
$$+ \dots + a_{n1} w_n, \; a_{22} w_1 + 2a_{22} w_2 + \dots + a_{2n} w_n$$
$$+ a_{12} w_2 + d_{32} w_3 + \dots + a_{n2} w_n, \dots)$$

b cont.

$a_{n1} w_1 + a_{n2} w_2 + \dots + 2a_{nn} w_n + a_{1n} w_1 + a_{2n} w_2 + \dots + a_{n-1\,n} w_{n-1})$

$= (2a_{11} w_1 + (a_{12} + a_{21}) w_2 + (a_{13} + a_{31}) w_3 + \dots + (a_{1n} + a_{1n}) w_n,\ (a_{12} + a_{21}) w_1 + 2a_{22} w_2 + (a_{23} + a_{32}) w_3 + \dots + (a_{2n} + a_{n2}) w_n, \dots, (a_{n1} + a_{1n}) w_1 + (a_{n2} + a_{2n}) w_2 + \dots + 2a_{nn} w_n)$

$$\nabla f(w) = \left(\begin{bmatrix} 2a_{11} & a_{12} + a_{21} & a_{13} + a_{31} & \dots & a_{1n} + a_{n1} \\ a_{21} + a_{12} & 2a_{22} & a_{23} + a_{32} & & a_{2n} + a_{n2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} + a_{1n} & a_{n2} + a_{2n} & a_{n3} + a_{3n} & \dots & 2a_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \right)^T$$

Therefore, $\nabla f(w) = \left([A + A^T] w \right)^T = w^T [A + A^T]$

Ositadinma Arimah
CS 4442B Assignment 1

c.

1c) let $f(w) = tr(Bww^T A)$

let the Hessian matrix H of f with respect to using the defn

let $f(w) = tr(Bww^T A)$ when A, $B \in \mathbb{R}$ are $n \times n$ matrices and $tr(A)$ is the squared matrixes' trace.

Because $tr(AB) = tr(BA)$

$tr\left(w^T \frac{Aw}{x}\right)$

$$tr[w_1, w_2, \dots, w_n] \begin{bmatrix} \sum_{i=j}^{n} a_{ij} w_j \\ \sum_{i=1}^{n} a_{ij} \dots w_j \\ \vdots \\ \sum_{i=1}^{n} a_{jj} w_j \end{bmatrix}$$

$$f(w) = \left(\frac{\partial f}{\partial w_1} + \frac{\partial f}{\partial w_2} + \dots \frac{\partial f}{\partial w_n} \right)$$

$$\nabla f(w) = [(A + A^T) w]^T$$

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

Ositadinma Arimah
CS 4442B Assignment 1

II.

1d) The sign of the eigen value determines the convexity along that direction.

(positive eigen values implies convex, negative implies concave, zero implies inflexion point)

$\therefore f(w) = tr(B w w^T A)$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

The eigen values of A =

$\lambda = \sqrt{2}, -\sqrt{2}$

The eigenvalues of B are

$\lambda = 3 \quad \lambda = 1$

$\therefore$ the function $f(w)$ is <u>NOT</u>

convex

Ositadinma Arimah
CS 4442B Assignment 1

a.

1e let $a(w) = w^T x$

let $b(w) = \log(\sigma(w))$

$\therefore \quad f(w) = b(a(w))$

Using the chain rule:

$$\nabla f(w) = \frac{df}{dw} = b'(a(w)) \cdot a'(w)$$

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

$$\Downarrow$$

$$a'(w) = x$$

$$b(a) = \ln(\sigma(a)) = \ln\left(\frac{1}{1+e^{-a}}\right)$$

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

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

$$b'(a) = \frac{-d(\ln(1+e^{-a}))}{da}$$

$$= \frac{-1}{e^{-a}+1} * \frac{d(e^{-a}+1)}{da}$$

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

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

e cont.b

Subbing $a(w)$ into $b(w)$ now we get:

$$F'(w) = b'(a(w)) \cdot a'(w)$$

$$= \frac{(e^{-w^T x})}{e^{-w^T x} + 1} \cdot x$$

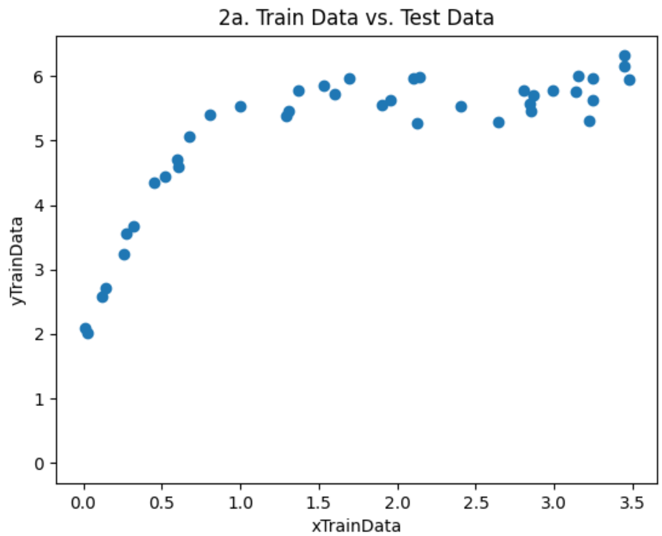
Therefore, the gradient of $f(w)$ is:

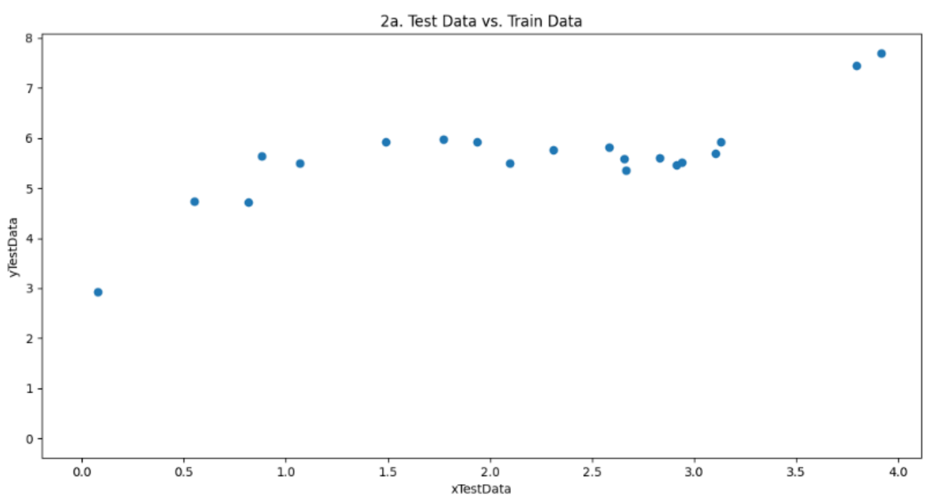
$$\nabla f_w = \frac{(e^{-w^T x})}{e^{-w^T x} + 1} \cdot x$$

Ositadinma Arimah
CS 4442B Assignment 1

III.

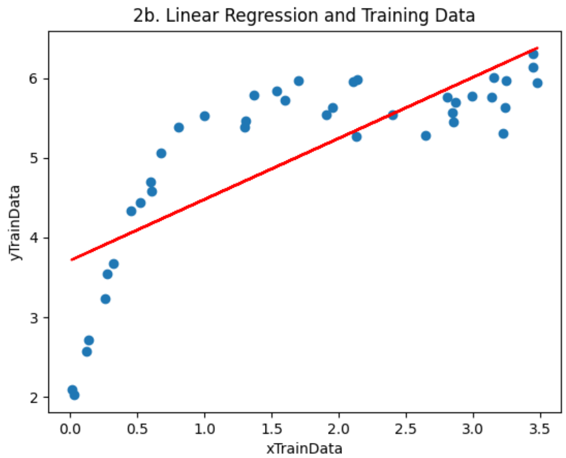
a.





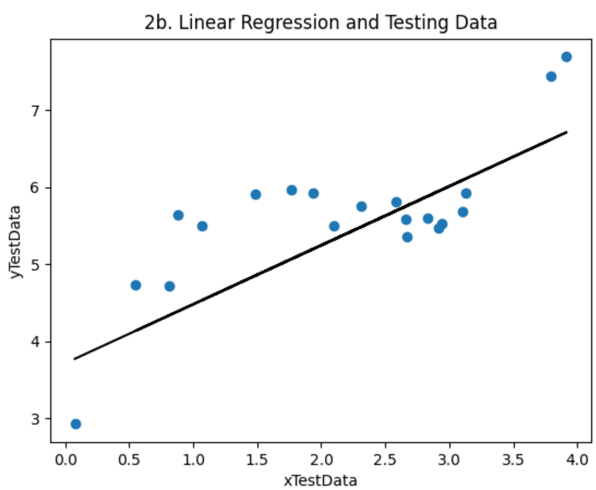
Ositadinma Arimah
CS 4442B Assignment 1

b.



The training error is: 0.5085888601660319

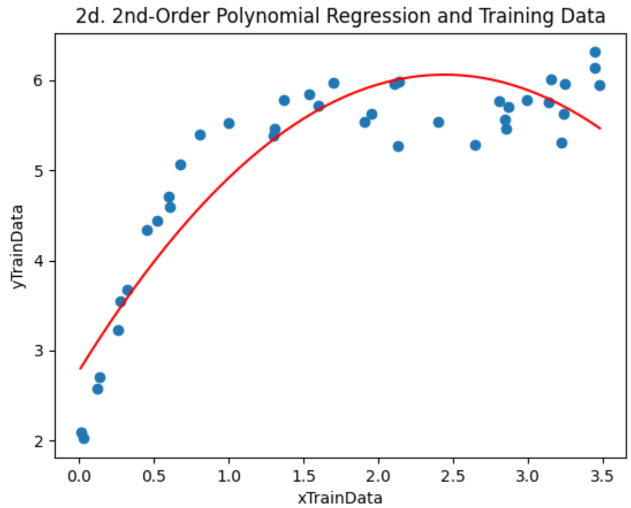
c.



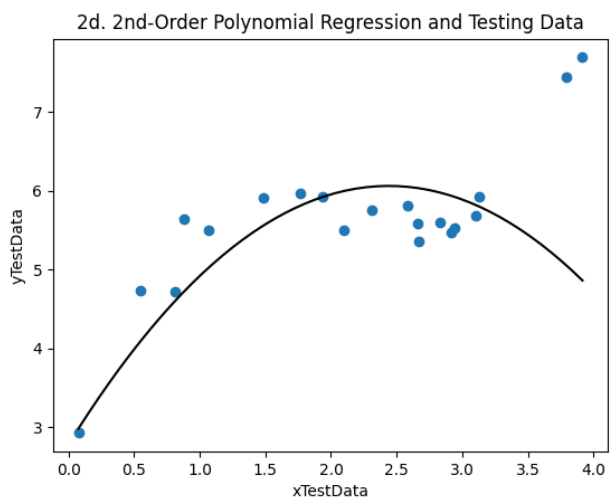
The test error is: 0.44391185790774834

Ositadinma Arimah
CS 4442B Assignment 1

d.



The training error is: 0.2009852319839666

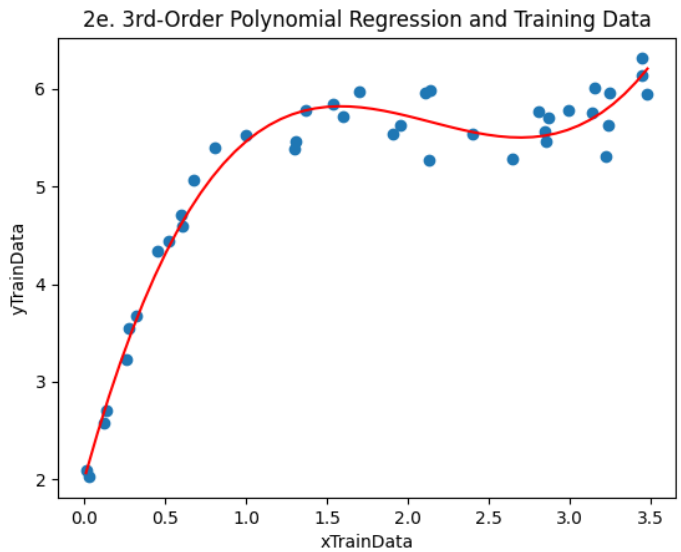


The test error is: 0.8532633206012309

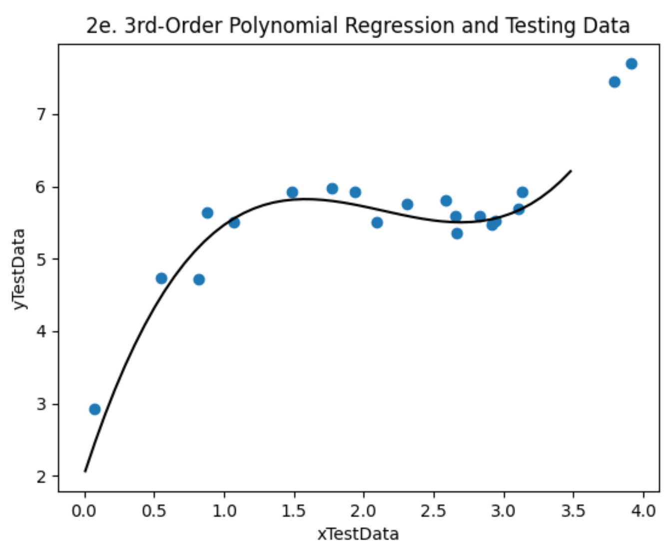
Therefore, the x^2 regression is a better fit than linear.

Ositadinma Arimah
CS 4442B Assignment 1

e.



The training error is: 0.03922874661114243

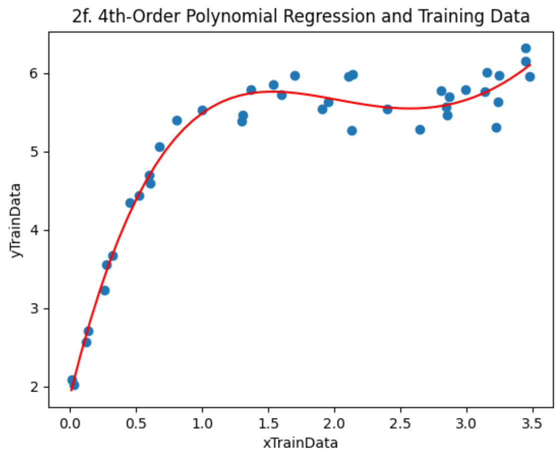


The test error is: 0.0564183300438727

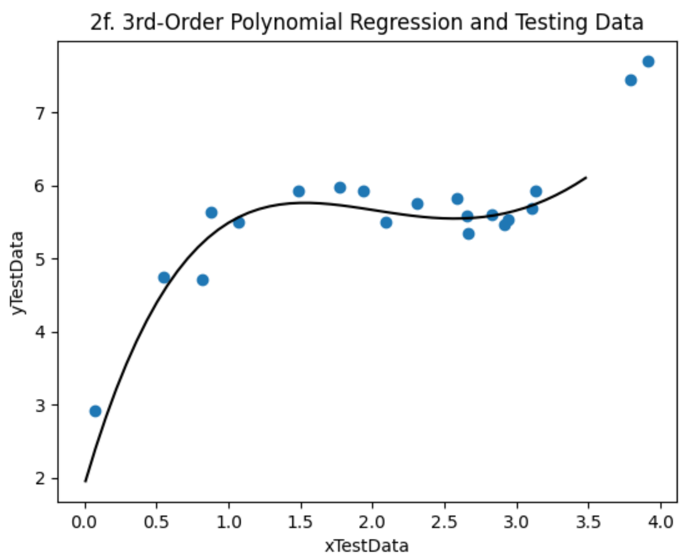
The x^3 regression is a better fit than linear and x^2 regression.

Ositadinma Arimah
CS 4442B Assignment 1

f.



The training error is: 0.03564470724439952

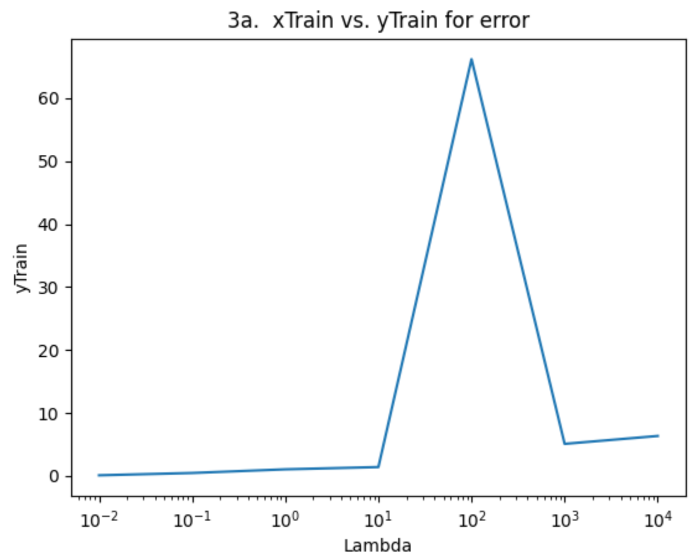


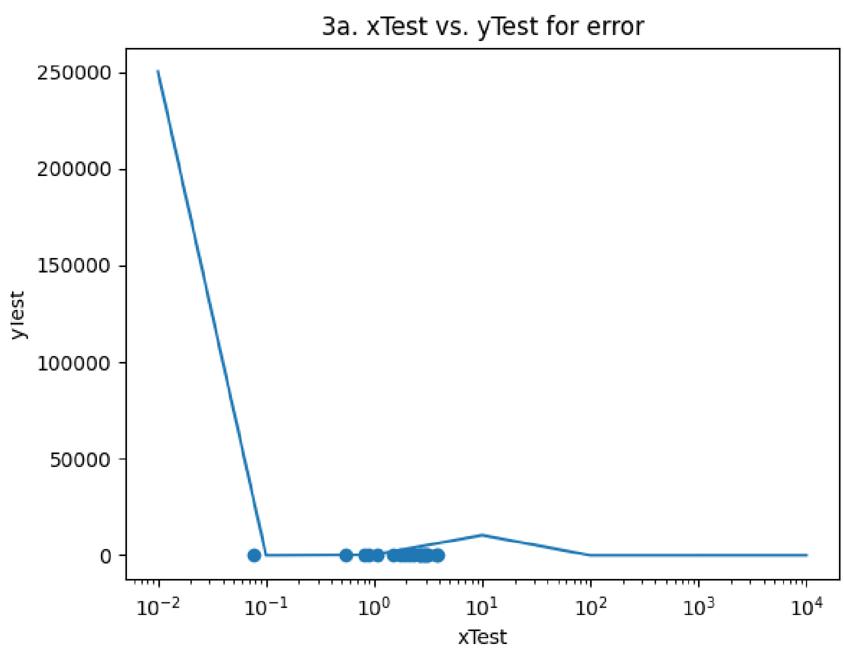
The test error is: 0.12722197193598808

Therefore, the x^4 regression has a better fit than linear regression. However, it does not have a better fit than x^2 or x^3 regression.

Ositadinma Arimah
CS 4442B Assignment 1

a.





The testing error is: 250212.87604839826

The testing error is: 0.542666017100274

The testing error is: 282.7298111725326

The testing error is: 10394.48064867868

Ositadinma Arimah
CS 4442B Assignment 1

The testing error is: 3.558108709085812

The testing error is: 5.708952334570834

The testing error is: 9.811652236909218

Therefore, the best fir for lambda is 0.1 as it has the lowest error for testing data.

b.