CMPT 419: Assigment 1 Fall 2018

Echo Liu

Linear Regression

1 Question 1

1. The nature of this distribution is of a multionomial distribution,
   as such its parameters are as follows
   $f(x_1, ..., x_k; n, p_1, ..., p_k) = Pr(X_1 = x_1 \ and ... and \ X_k = x_k) = \frac{n!}{x_1!...x_k!} * p_1 \times ... \times p_k$
   where each p_k can be thought of as μ_k and $\sum_{n=1}^{k} x_i = n$

2. Values of each μ_i is $\frac{1}{3}$

3. Let Ω be our sample space, let B be our event space and
   P be our probability function and j be an index that denotes
   that party j is elected, then:
   Ω :
   $$\{[0, 0, 1], [1, 0, 0], [0, 1, 0]\}$$
   where we compute
   the permutations of [[0,0,1],[1,0,0],[0,1,0]], [[1,0,0],[0,1,0],[0,0,1]]

   B:
   $$\{0, 1, 2\}$$
   in otherwords each element in this set is an index set into set Ω where each index denotes that a particular party is elected for example if I computed Ω_b where b is in set B then I am essentially indexing the bth element of Omega e.g Ω_1=[1,0,0] then in index j, we get our outcome if and only if j=b i.e if j=b, then P evaluates to 1, 0 otherwise, where j is an index into the bth element In otherwords
   $$P(j) = \begin{cases} 1 & \text{if b=j} \\ 0 & \text{otherwise} \end{cases}$$

4. For this question, imagine oneself as an auditor, auditing election practices. As such if a politician wanted to rig the election they'd hire a partison auditor. As such the components of the alpha vector would have components $[\alpha_1, \alpha_2, ..., \alpha_i] where for party k, we use a one hot encoding scheme and subsitute in our Dirshlet Distribution as appropiate.$

5. We assume that when a given party is elected, that this distribution takes on a Dirchlet distribution. Moreover, we take all the alpha parameters and set them to 1. In otherwords we are encoding an uniform Dirshlet prior, with a one hot encoding scheme about the alpha vector for the Green party.

6. Assuming each party has equal probability of winning, we take the Dirschlet distribution and intergrate as appropiate.

2 Question 2

Basic idea: Assign a subscript i, to each beta to form β_i
$Hence \quad we \quad do \quad the \quad following$

$p(\vec{t}|\vec{w}^T\phi(\vec{x_n}), \beta_n^{-1}) = \prod_{n=1}^{N} \frac{\sqrt{\beta_n}}{\sqrt{2\pi}} exp((-\frac{\beta_n}{2}(t_n - \vec{w}^T\phi(\vec{x_n}))))$
Then we change the product to a summation
$= \sum_{n=1}^{N} ln(\frac{\sqrt{\beta_n}}{\sqrt{2\pi}} exp((-\frac{\beta_n}{2}(t_n - \vec{w}^T\phi(\vec{x_n})))))$
We approach each variable as a joint density function so we obtain the following:

$= \sum_{n=1}^{N} ln(2\pi)^{-\frac{1}{2}} + \sum_{n=1}^{N} ln(\beta_n^{-\frac{1}{2}}) + \sum_{n=1}^{N}((\frac{1}{2}\beta_n(t_n - \vec{w}^T\phi(\vec{x_n})))$
One can then simplify the expression further by moving the 1/2 to the front
Our sum of squares error term is the third term

3 Question 3

Part 1: The answer is no : the reason being is that the validation set could possibly provide the model with some insights into the actual model

The answer is no , your data could happen to be modeled as a straight line and thus increasing the polynomial degree will increase the training error

The answer is yes : by definition regularizion penalizes model complexity

4 Question 4

We are essentially trying to find the gradient of the elastic net regression function

Thus our error function goes as follows
$E_D(\vec{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \vec{w}^T\phi(\vec{x_n})\} + \sum_{i=1}^{J_1}\lambda_i|w_i| + \sum_{j=1}^{J_2}\frac{\lambda_j}{2}w_j^2$
Therefore our gradient goes as follows, our approach is that we use the matrix form of the quadratic i.e think of x^2 and $\vec{x}^T\vec{x}$ as logically equivalent, thus we obtain the following for the gradient
$\nabla E_D(\vec{w}) = -\sum_{n=1}^{N}\{t_n - \vec{w}^T\phi(\vec{x}_n)\}\vec{\phi}^T(\vec{x}_n) + \frac{1}{2}\sum_{i=1}^{J_1}\lambda_i\frac{\vec{w}}{|w_i|} + \sum_{j=1}^{J_2}\lambda_j|w_j|$

5 Question 5

1. Responses to section section "Getting Started"
   The country that had the highest child mortality rate in 1990 is Libera
   with value of 160.8
   The country that had the highest child mortality rate in 2011 is Sierra Leone
   with value 119.2
   The API handles the NaN values by filling them with random entries, in practice
   this might not be the correct thing to do.

2. Response to section "Polynomial Regression"

[b]0.4

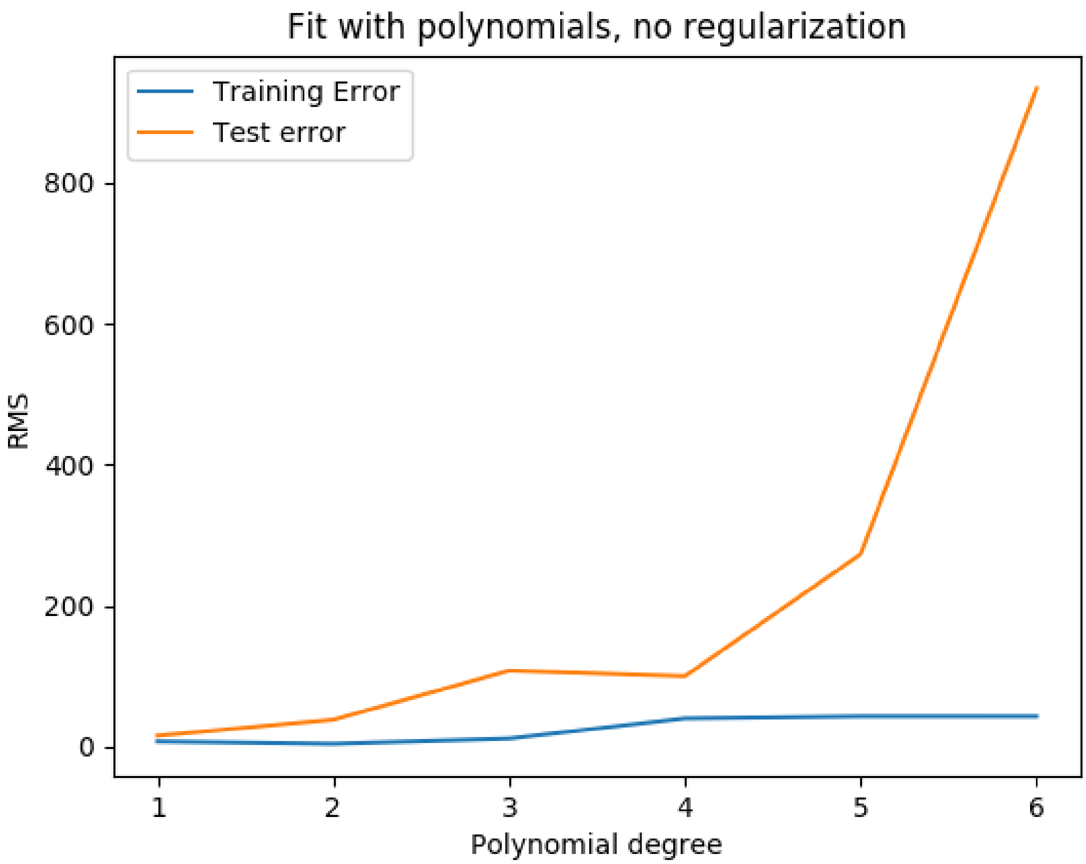


Figure 1: Non normalized features, we plot the error in RMS versus polynomial degree as one can see the testing error is greater than training error

6 Question 5 Continued

[b]0.4

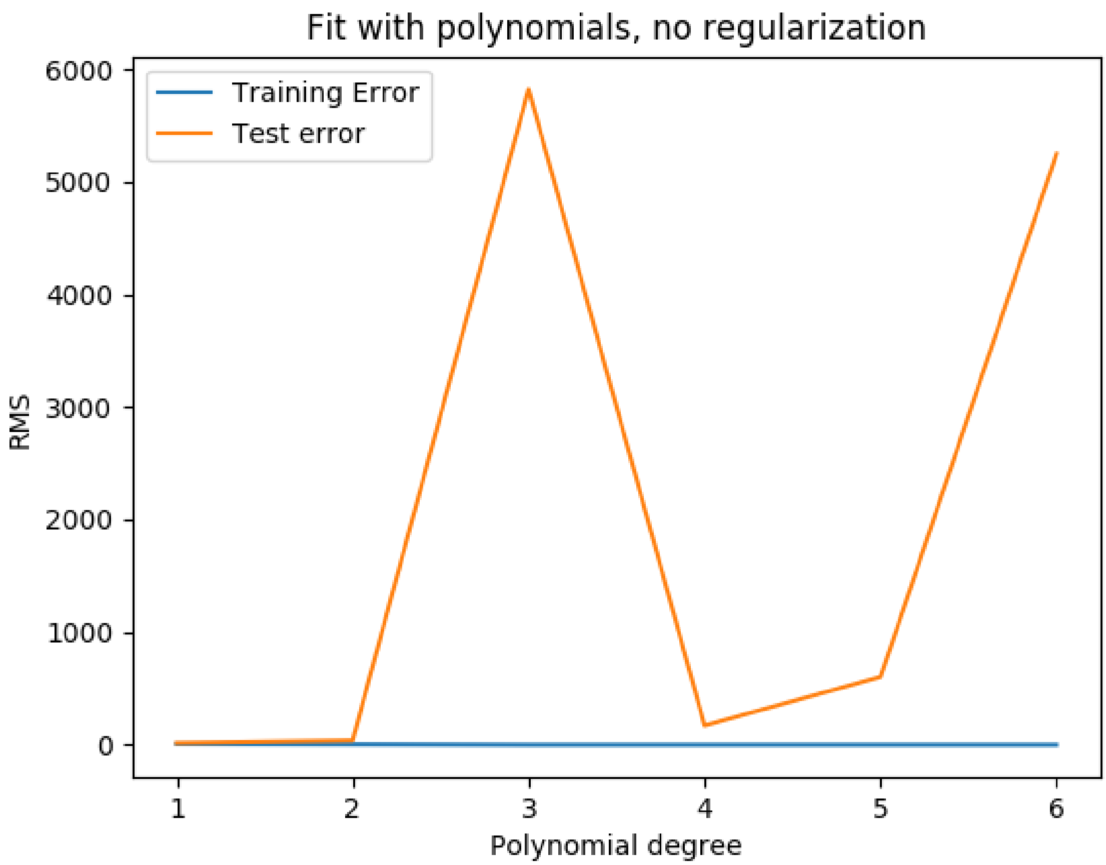


Figure 2: We run our program again, normalizing our features

7 Question 5 Continued

[b]0.4

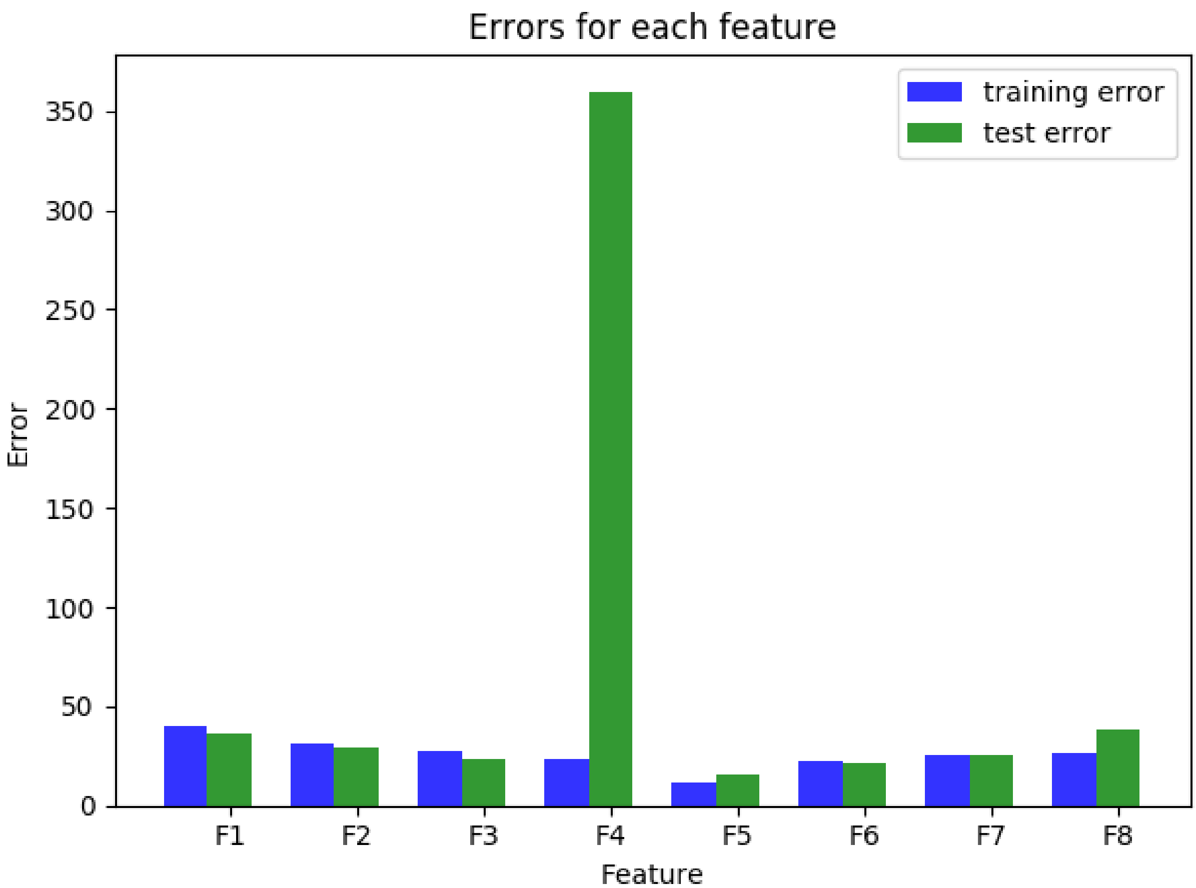


Figure 3: As one can see, there is a spike in error in one of our features, suggesting a polynomial fit might not best describe out data

8 Question 5 Continued

[b]0.4

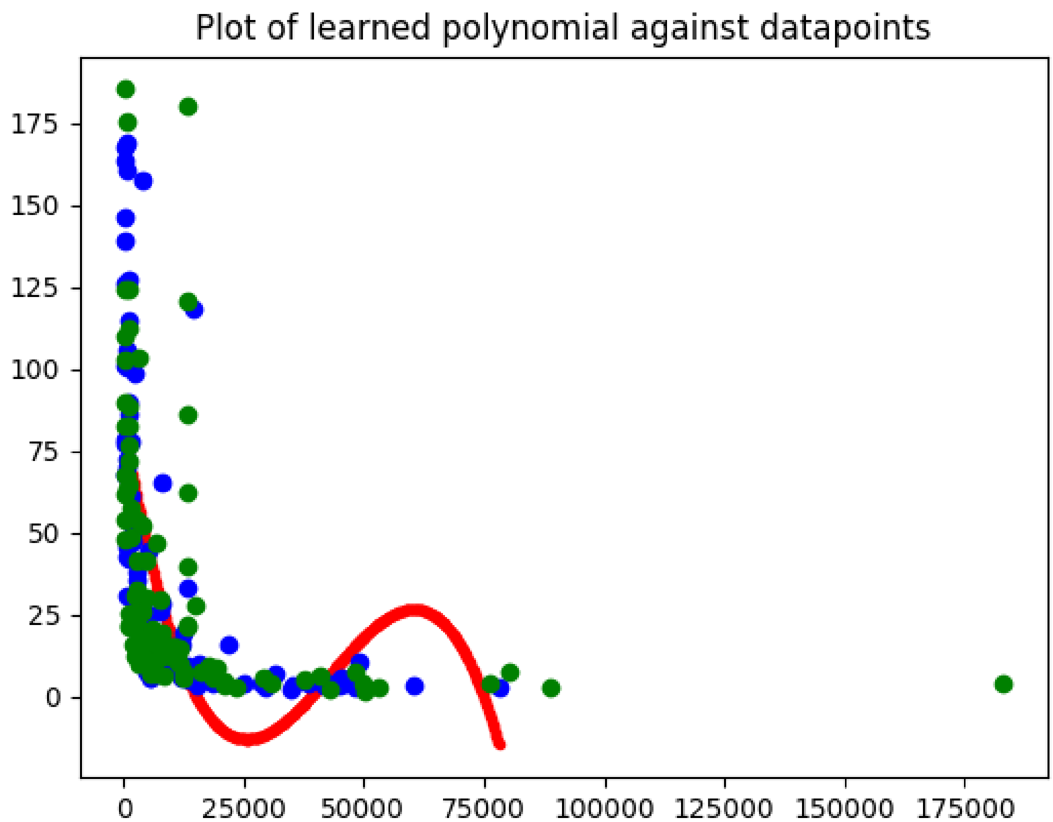


Figure 4: Our curve fitted against our training and test points with our GNI feature, as one can see the polynomial does not best fit our data

9 Question 5 Continued

[b]0.4

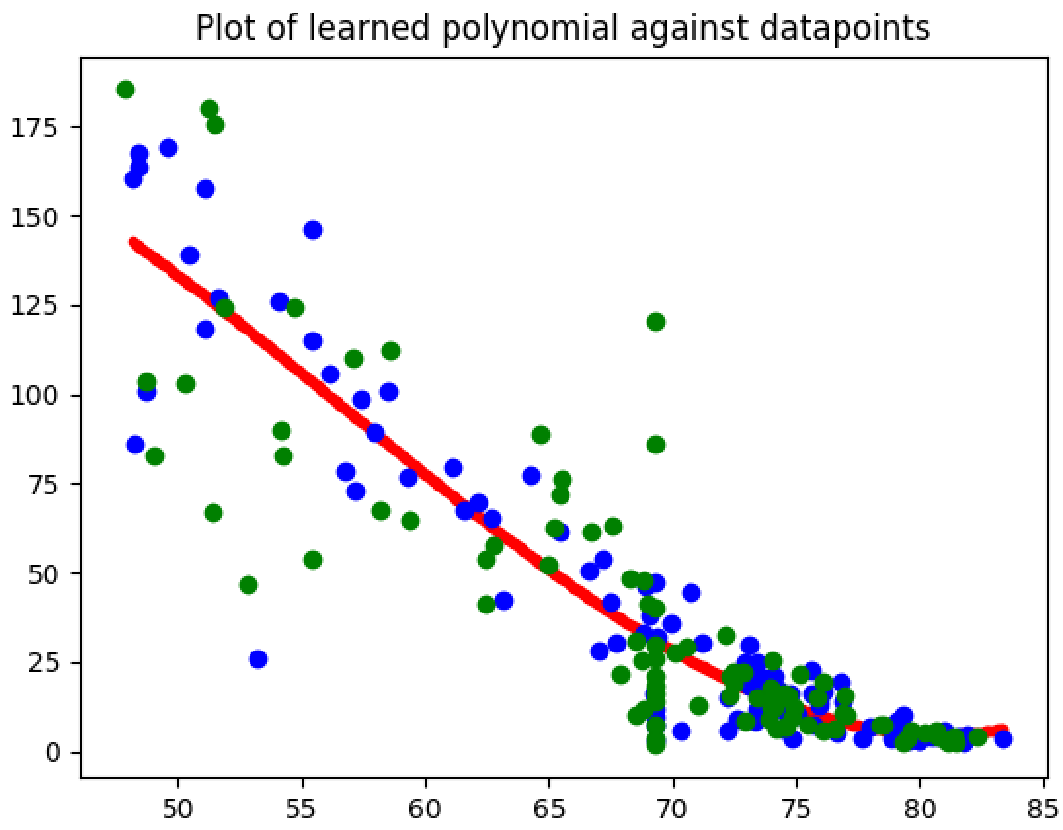
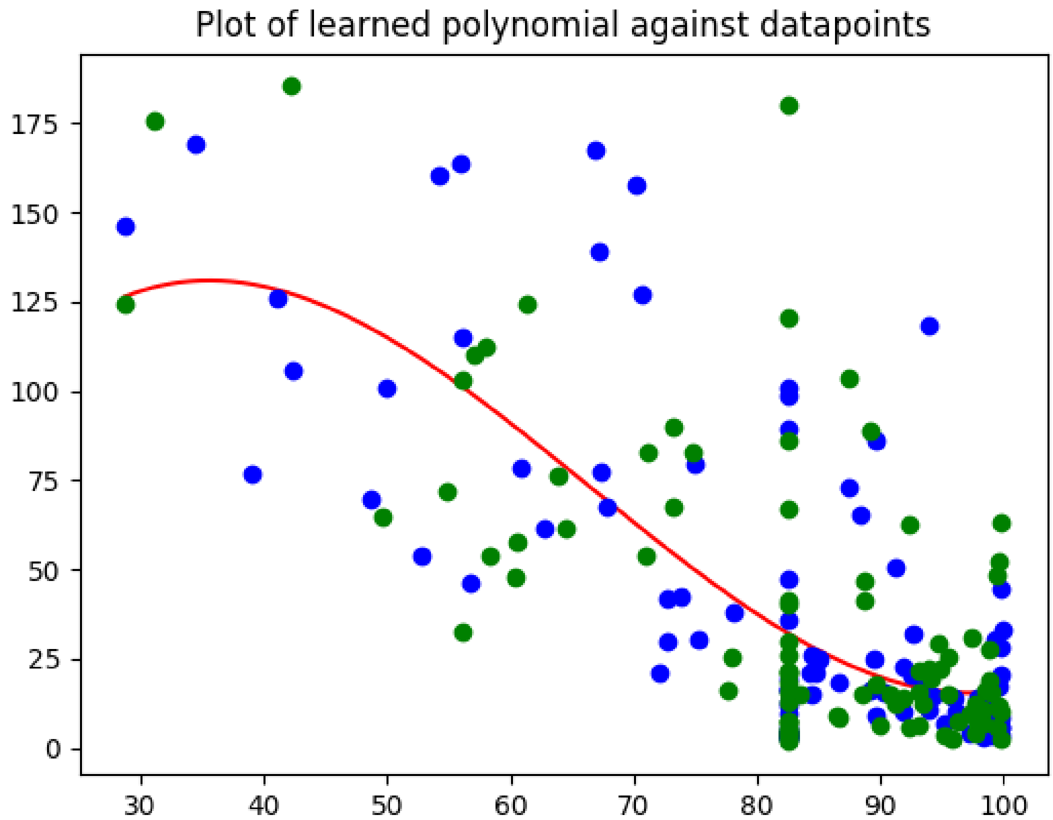


Figure 5: Our curve fitted against our training and test points with our Life Expectancy feature,as one can see the data isn't best represented by a polynomial of degree 3

10 Question 5 Continued

[b]0.4



Our curve fitted against our training and test points with our Literacy Expectancy feature, as one can see a deg 3 polynomial does not best fit our data

11 Question 5 Continued

Response to section "ReLU Basis Function"

[b]0.4

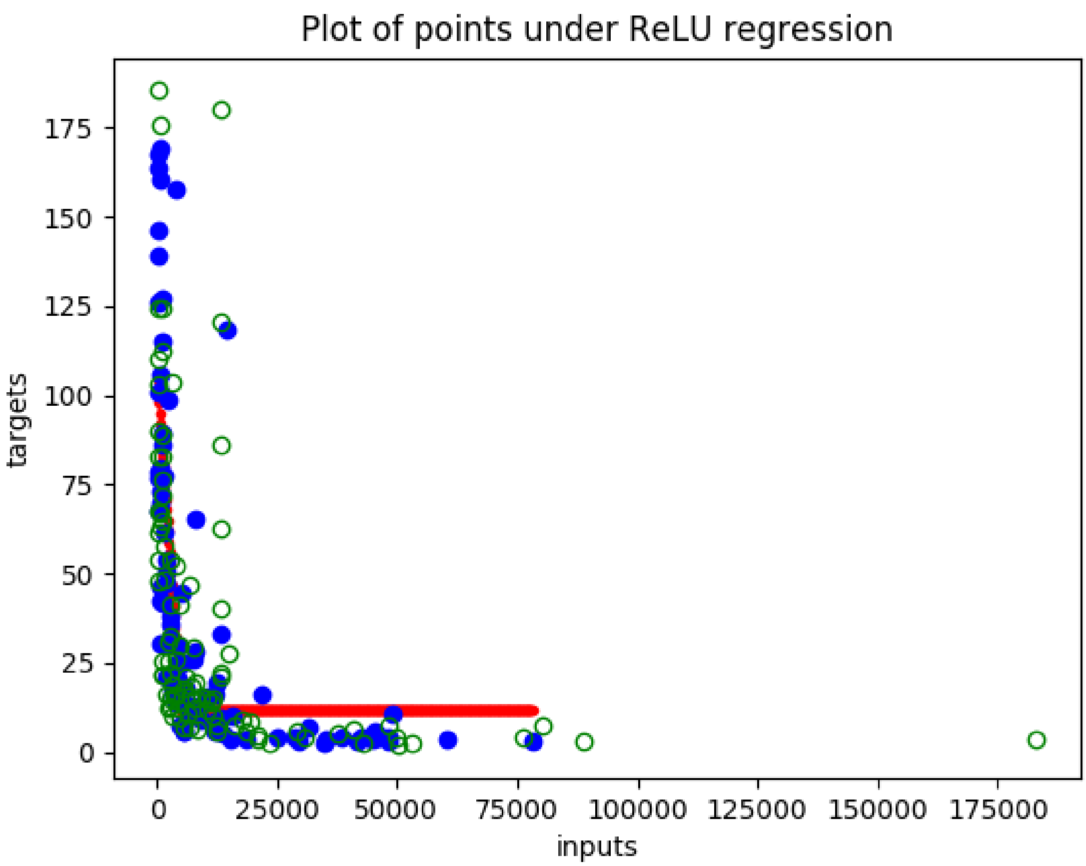


Figure 6: Our curve fitted against our data [test and training] points, with training and testing error resp being 20.59 and 24.199

12 Question 5 End

Response to section "Polynomial Regression, Regularization"

[b]0.4

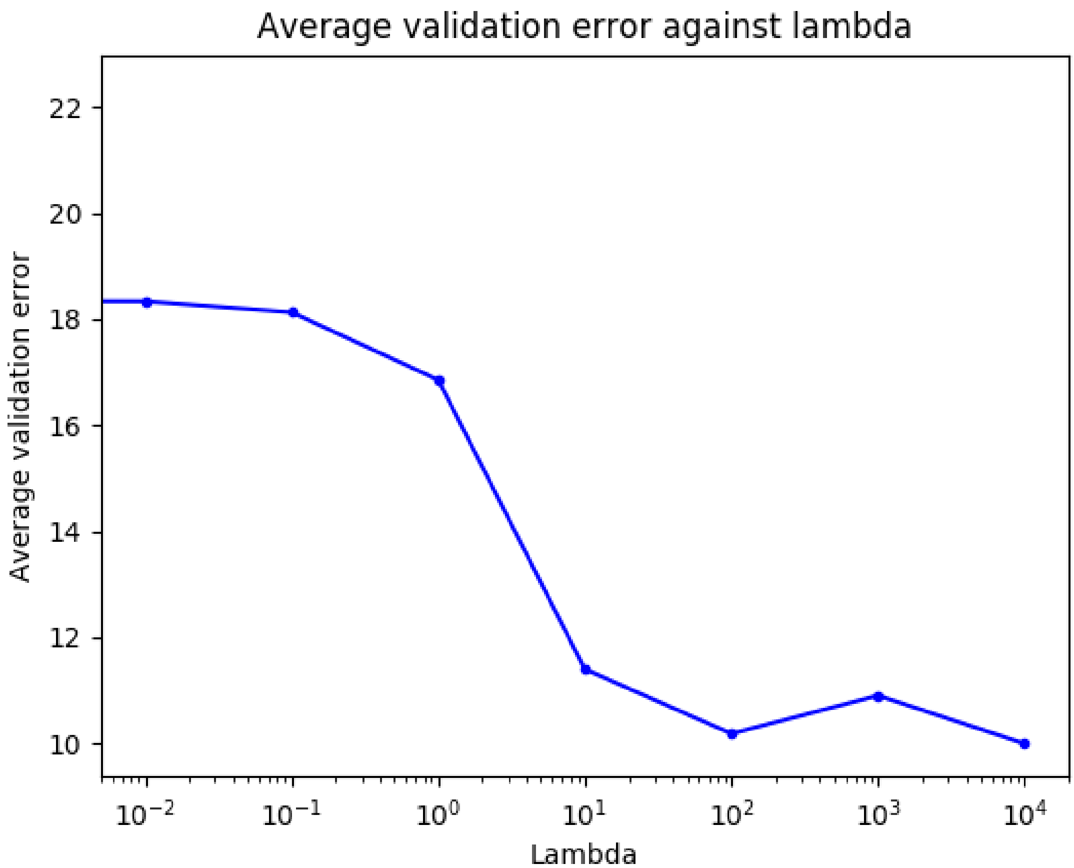


Figure 7: As one can see in the following diagram, the best value for lambda is 100 or ten squared and tied with 10000 or 10 to the four