Assignment-1 ELL409

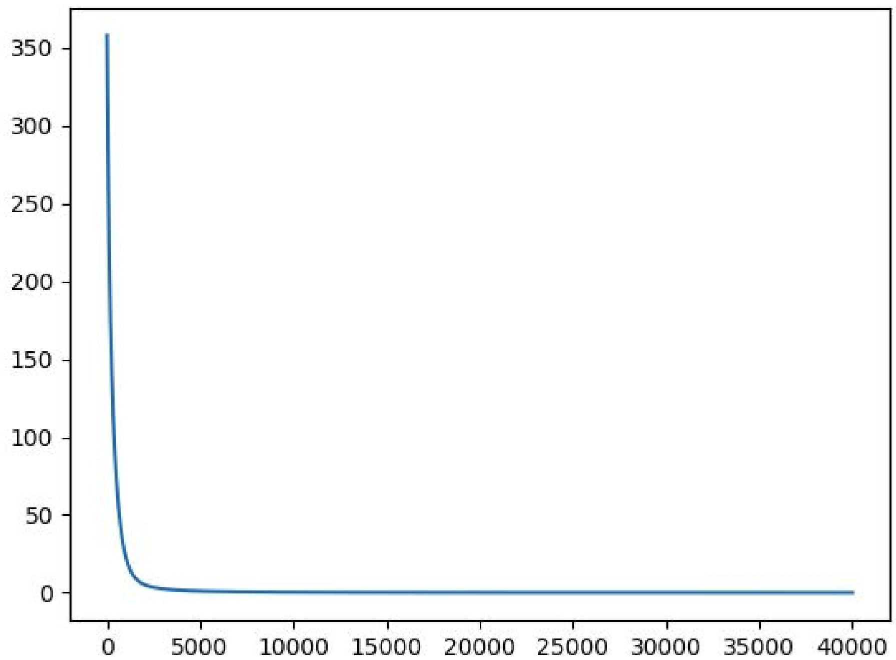
Submitted By:

Ritik Agrawal (2017EE10482)

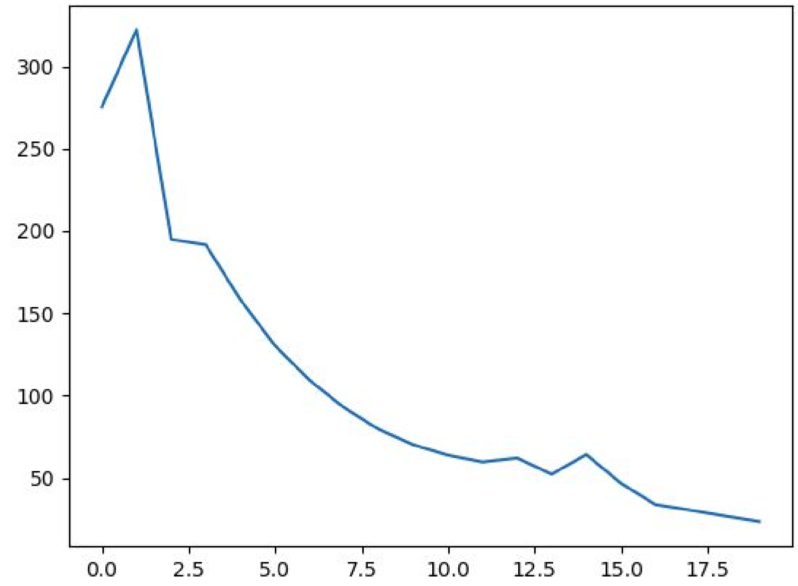
Part-1

1) For 20 data points:

    a) Error vs iteration-

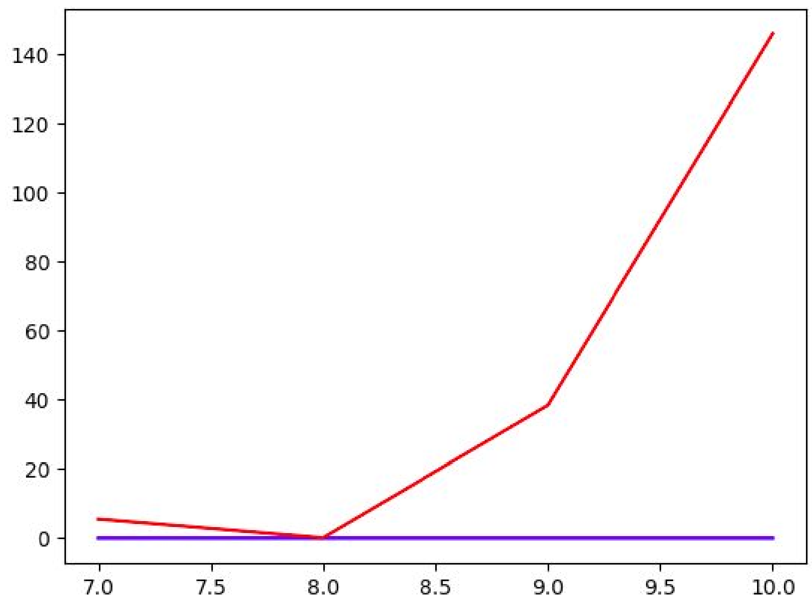


b) Error vs batch size-



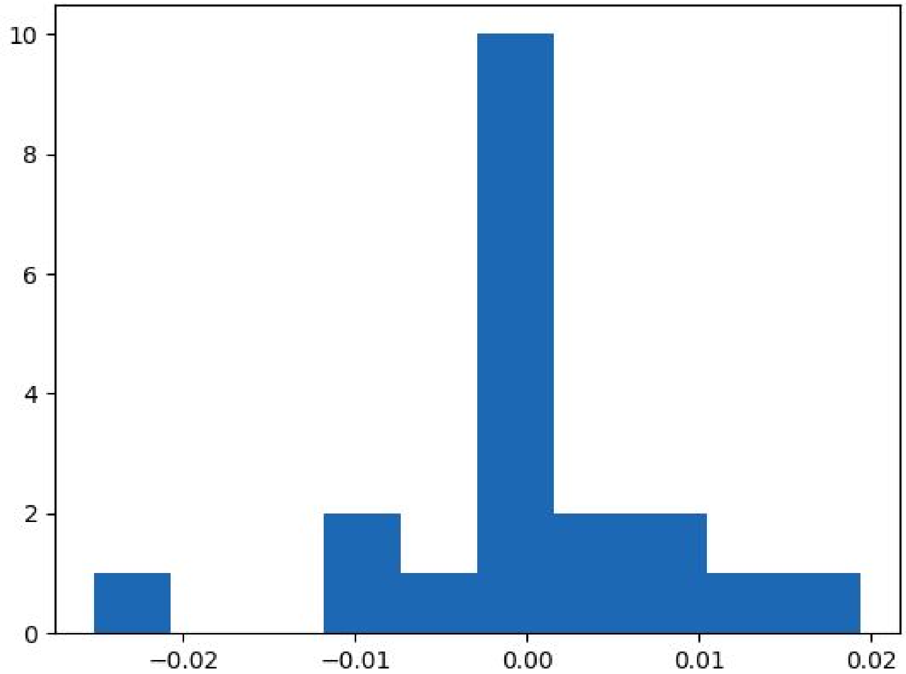
Note: Error decreases as we increase the batch size, hence full batch is desirable for the gradient descent.

c) Validation Plot for degree of the polynomial-

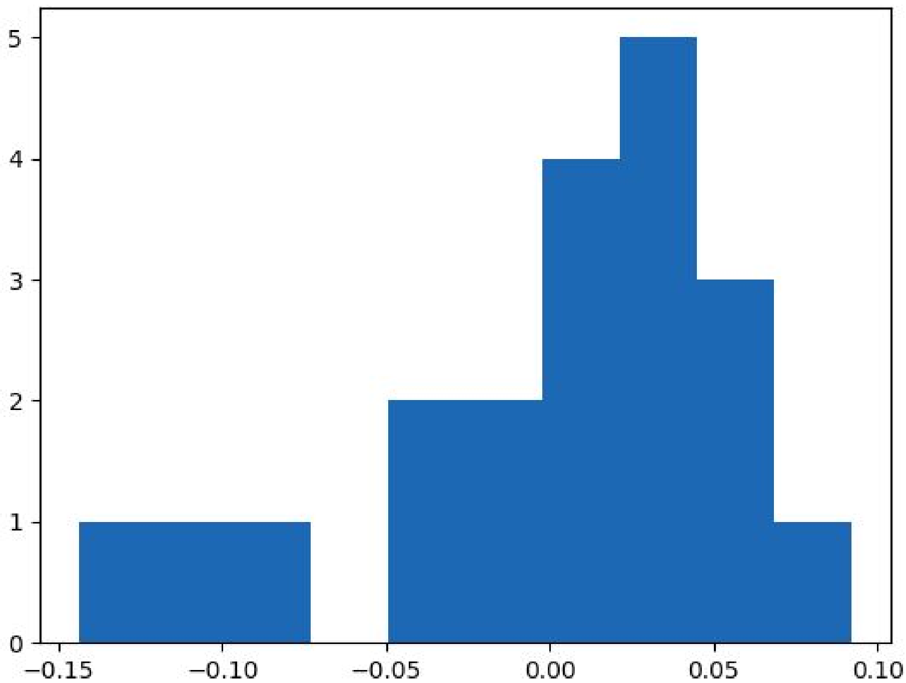


Note: M= 8 is the desirable degree of the polynomial

d) Residual error plot-

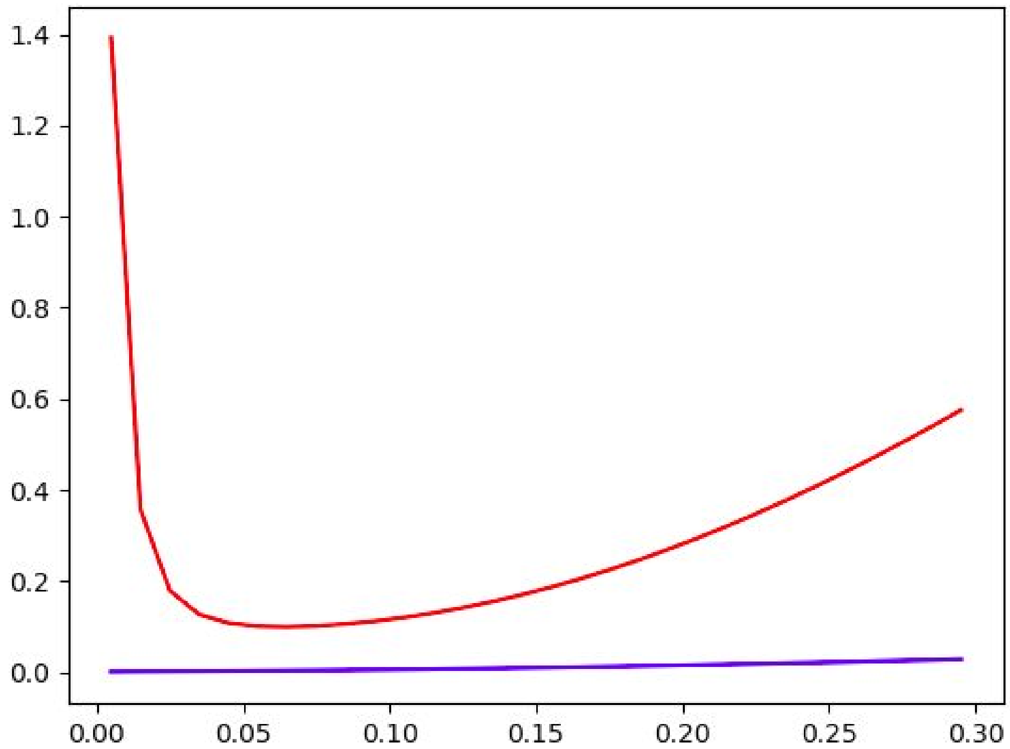


Variance- 7.812984729270083e-05 (Moore Penrose)



Variance- 0.0032585990649981753 (Gradient Descent)

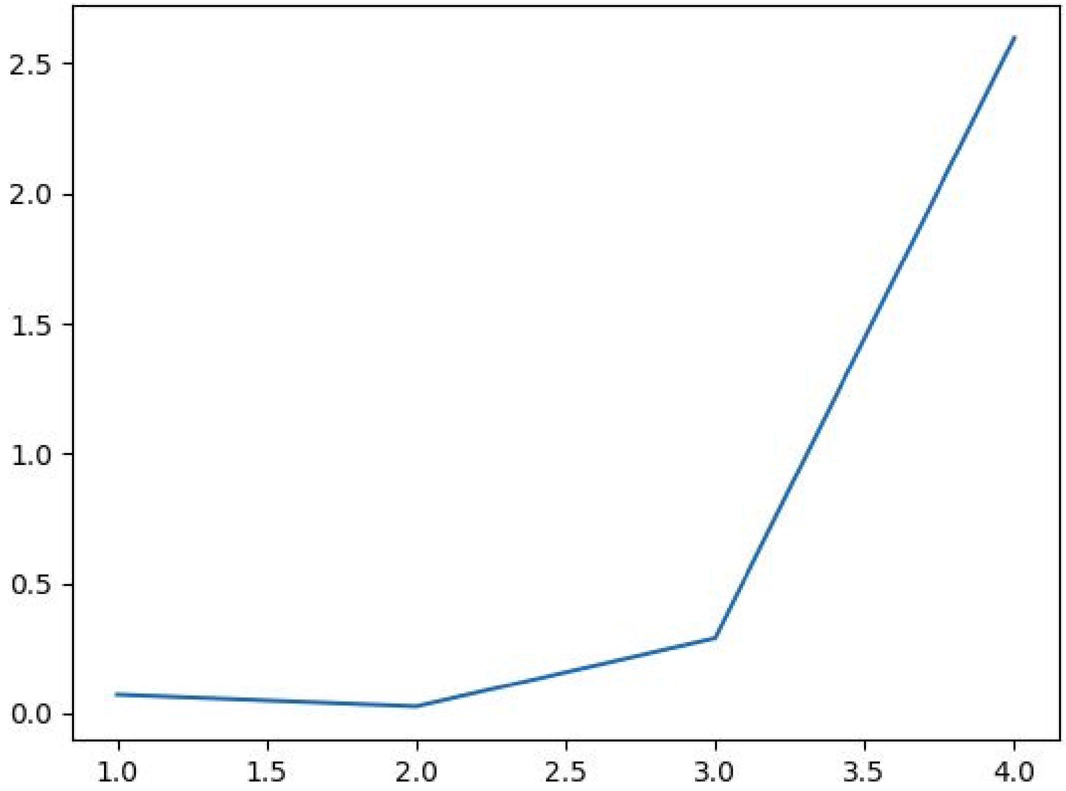
e) <u>K-fold validation plot for regularisation coefficient-</u>



Note: Regularisation constant comes around 0.065 from the cross-validation curve

w= [5.92 0.043 0.23 -0.16 -0.08 0.092 -0.067 0.113 -0.05]

f) <u>Lq error plot-</u>



Final error for different values of q-

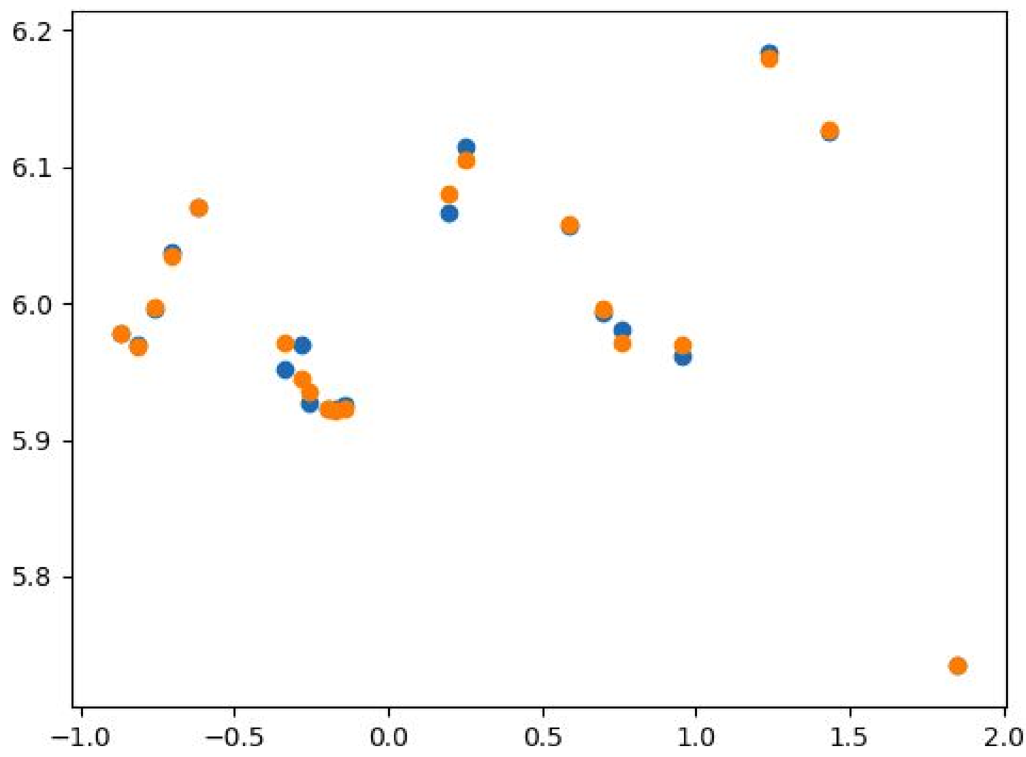
i) q=1, a=1.5*1e-4, e=0.072
ii) q=2, a=3.6*1e-5, e=0.027
iii) q=3, a=3.6*1e-6, e=0.29
iv) q=4, a=4.7*1e-7, e=2.6
v) q=5, a=6*1e-8,   e=20.37

Hence, q=2, which is the squared loss function is the ideal fit for our model.

Results-

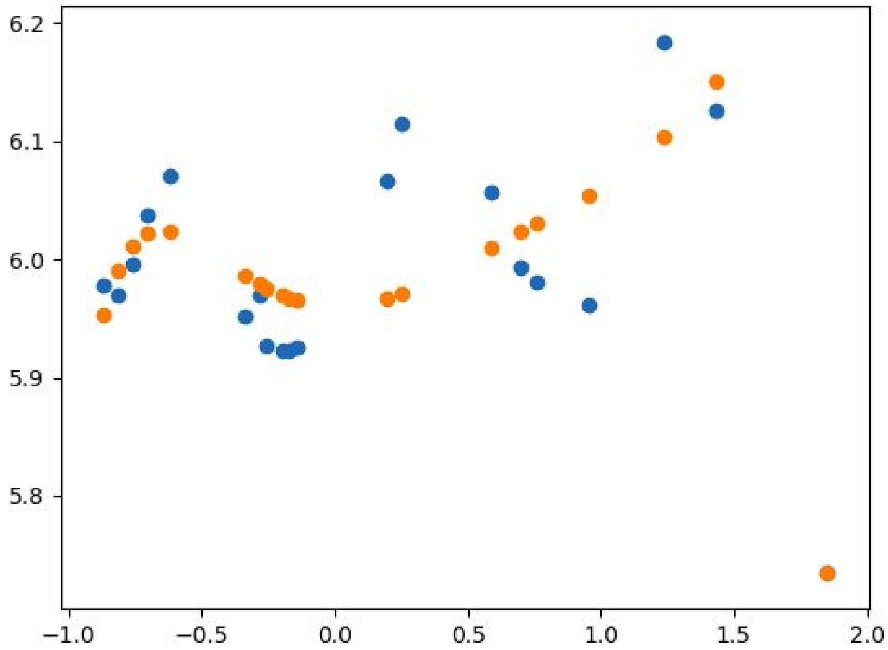
a) Moore-Penrose Inverse Matrix-

Final Error- 0.0007812984729270008
w= [5.97 0.51 0.94 -3.09 -2.18 5.67 0.42 -3.32 1.072]



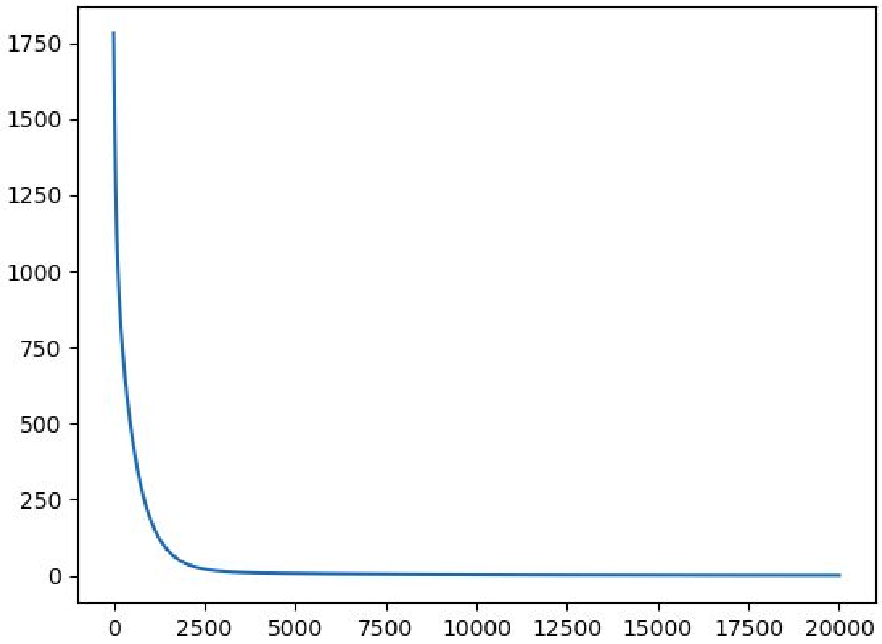
b) Gradient Descent-

No. of Iterations- 80000
Final Error- 0.03258599064998175
Learning Rate- 7.2*1e-5
w= [5.96 -5.513e-03 2.19e-01 -4.92e-02
 -1.38e-01 6e-02 -6.57e-02 1.37e-01 -5.74e-02]

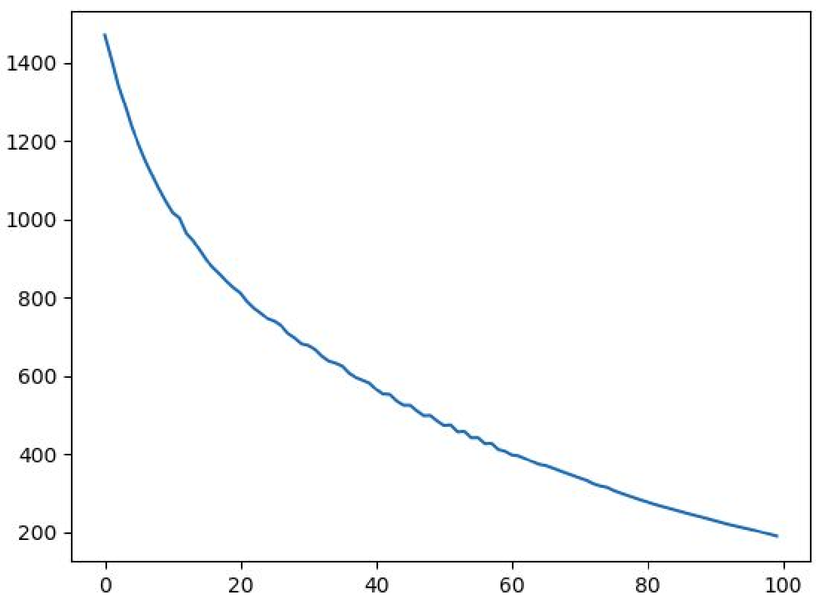


2) For 100 data points:

a) Error vs iteration

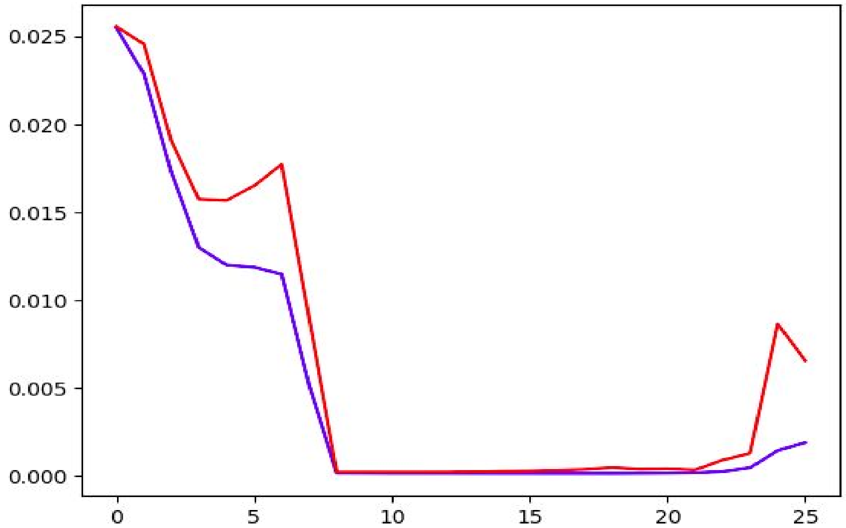


b) <u>Error vs batch size-</u>



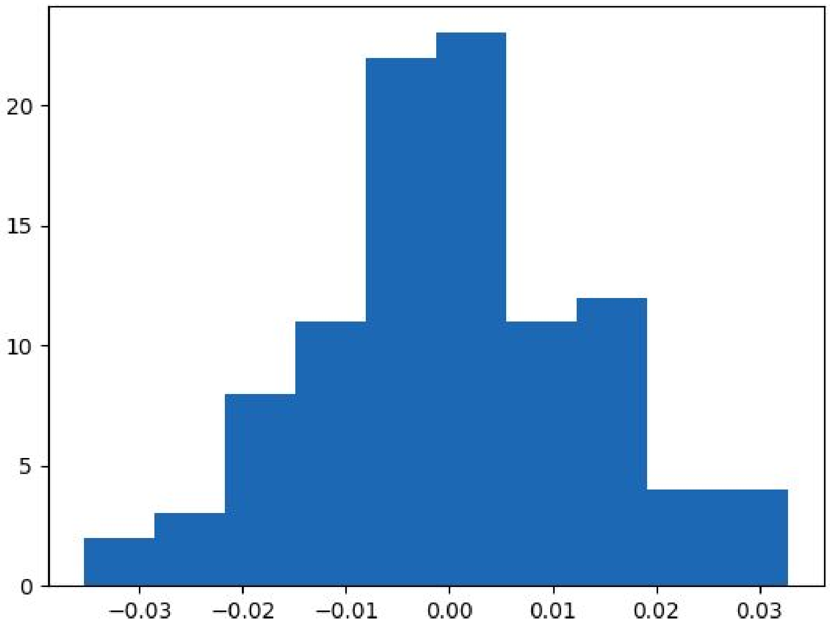
Note: Error decreases as we increase the batch size, hence full batch is desirable for the gradient descent.

c) <u>Validation Plot for degree of the polynomial-</u>

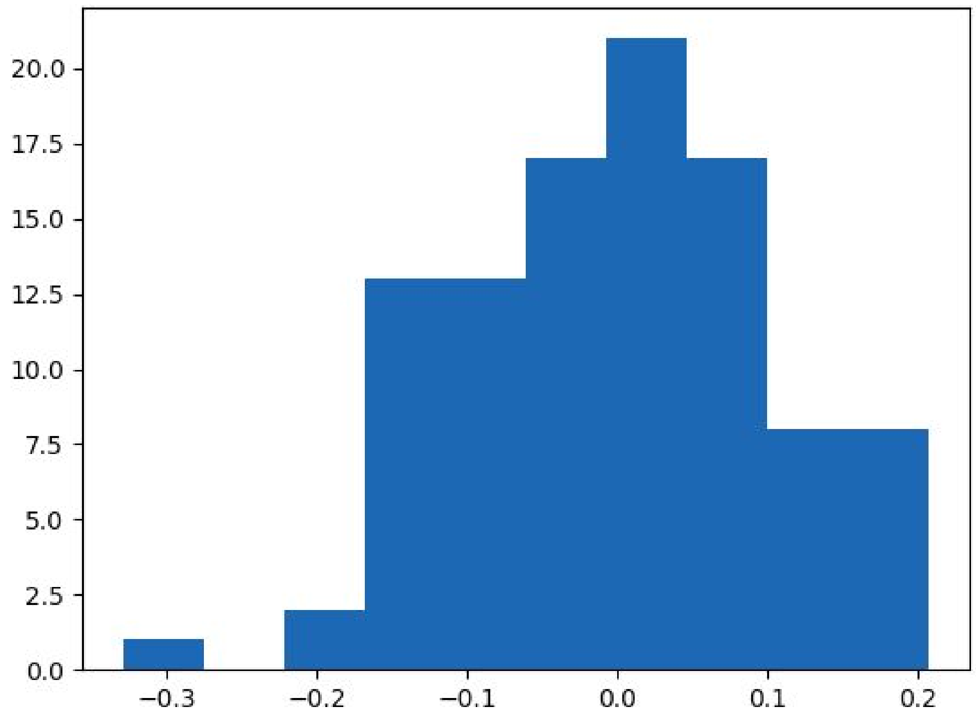


Note: M= 8 is the desirable degree of the polynomial

d) <u>Residual error plot-</u>

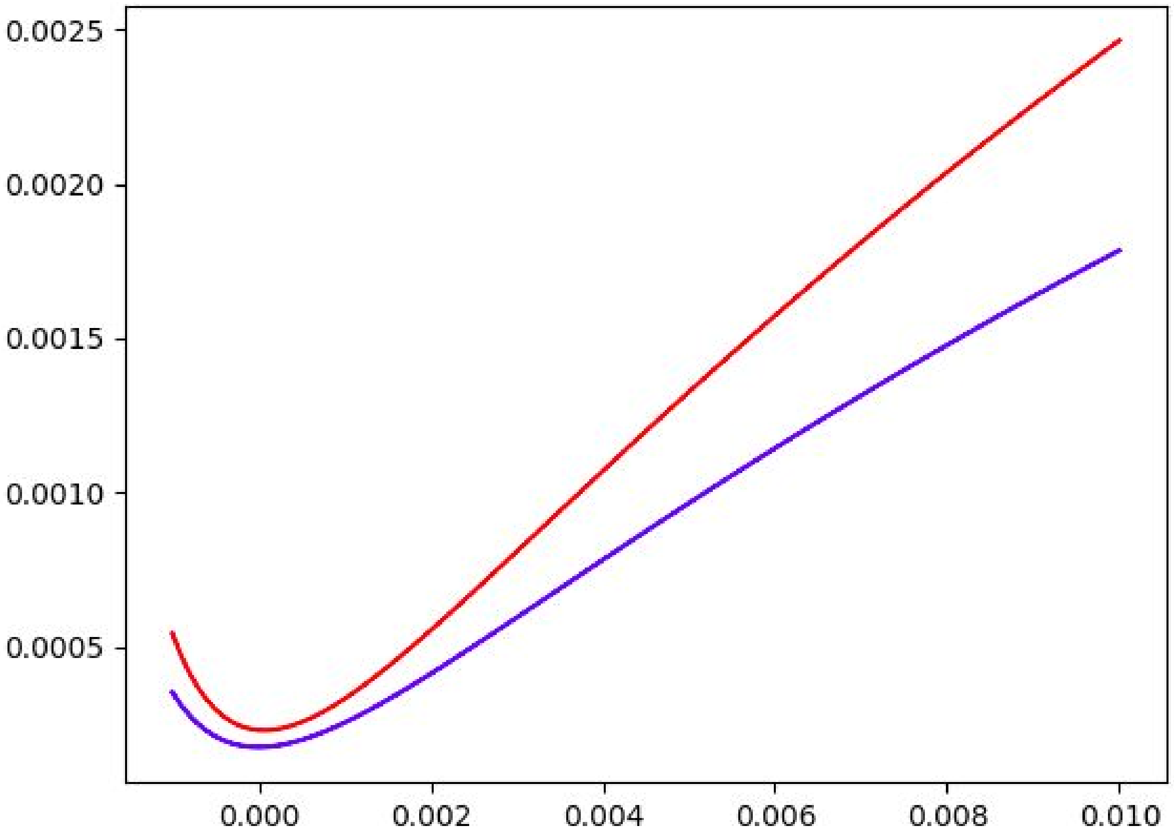


Variance- 0.00017987257191094536 (Moore-Penrose)



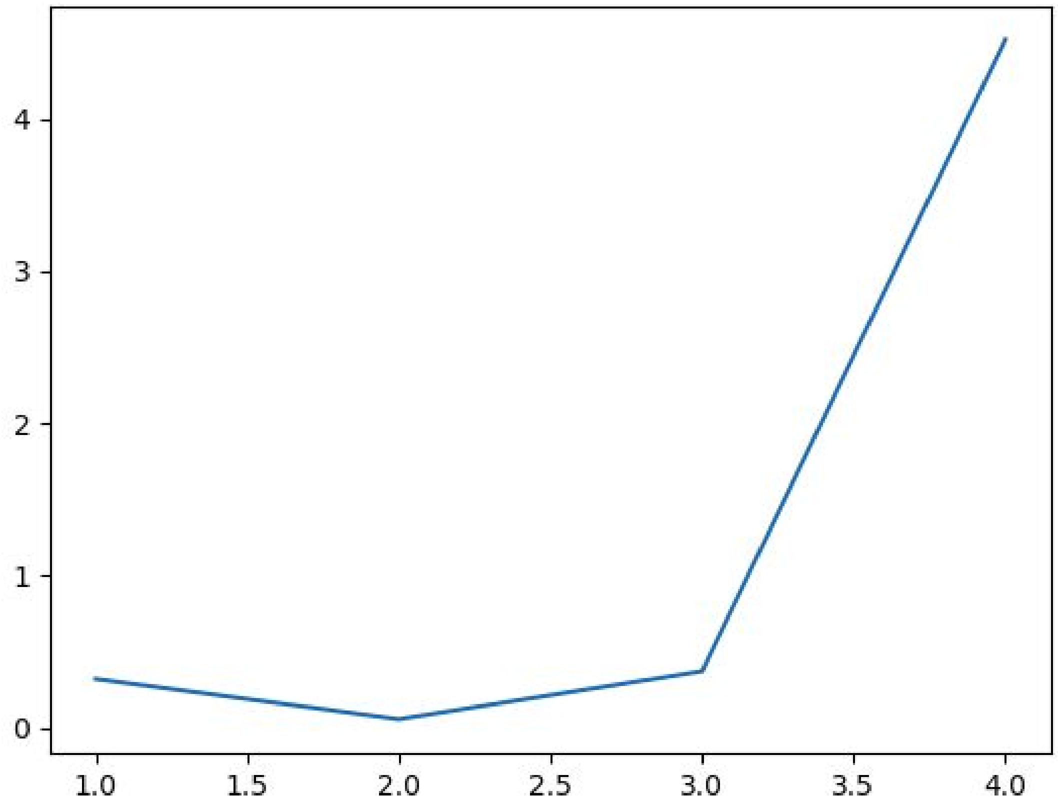
Variance- 0.010288513182148695 (Gradient Descent)

e) <u>K-fold validation plot for regularisation coefficient-</u>



Note: Regularisation constant comes around 0 from the cross-validation curve

f) <u>Lq error Plot-</u>



Final error for different values of q-

i) q=1, a=1.25*1e-5, e=0.32
ii) q=2, a=6.25*1e-6, e=0.055
iii) q=3, a=5.9*1e-7, e=0.37
iv) q=4, a=7.1*1e-8, e=4.52
v) q=5, a=9.2*1e-9, e=34.51

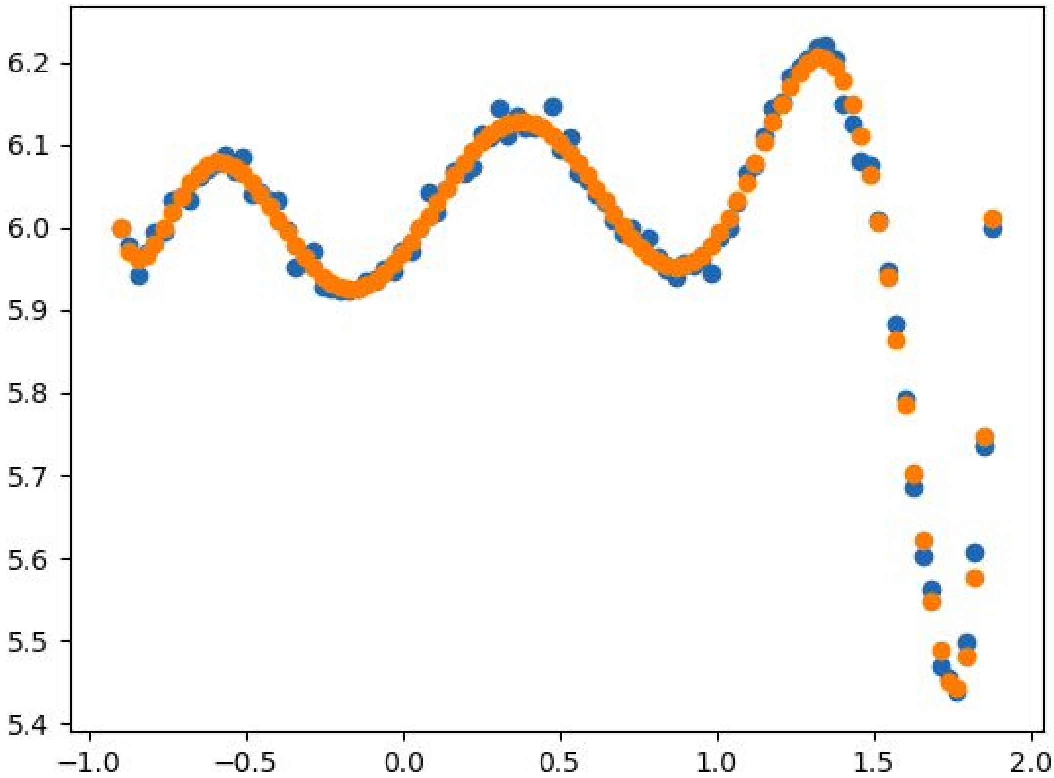
Hence, q=2, which is the squared loss function is the ideal fit for our model

Results-

a) Moore-Penrose Inverse Matrix-

Final error- 0.008993628595547264

w= [5.97 0.49 0.97 -2.97 -2.26 5.47 0.49 -3.19 1.02]



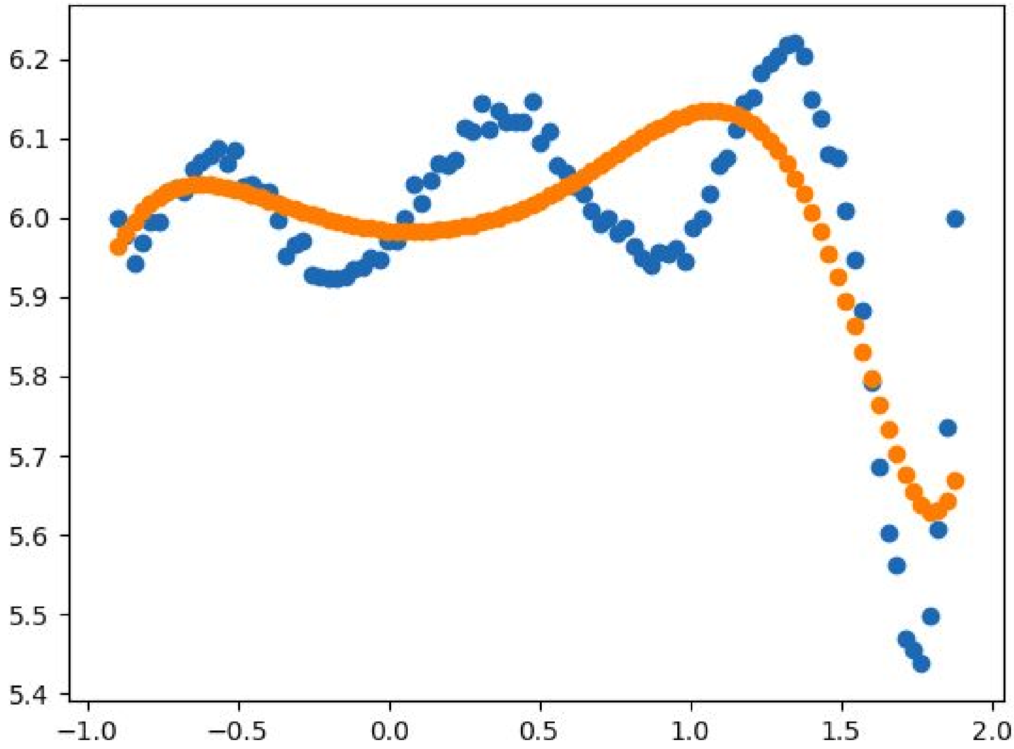
b) Gradient Descent-

No. of Iterations- 80000

Final error-0.5144292144595836

Learning rate= 1.25*1e-5

w= [5.98 -0.03 0.198 0.03 -0.05 0.18 -0.17 -0.05 0.05]

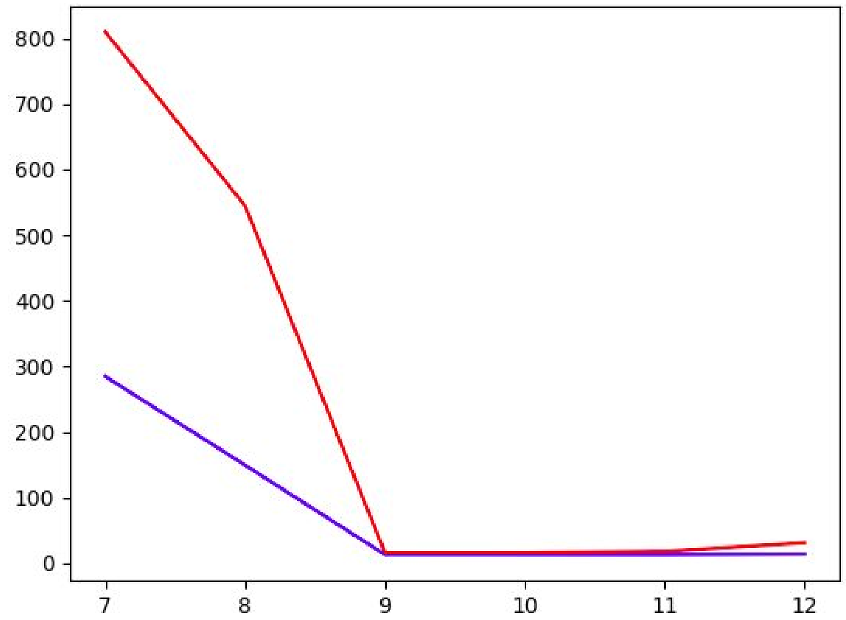


Discussion-

1) The histogram for the residual error in case of 20 data points didn't come out exactly to be gaussian as in the case of 100 data points due to lack of data.
2) The regularisation parameter came out to be non zero for 20 data points as opposed to 100 data points to reduce the model complexity which was not required for the latter dataset.
3) However, the trends were similar for both the datasets and optimal M=8 for both of them. Below M=8 underfitting was observed and above it, overfitting was observed.
4) The error in gradient descent and moore penrose were both relatively higher 100 data points. This can due to faster convergence in case of 20 data points.
5) Convergence criterion- On observing the error vs iteration plot, after 80000 iterations, the error was decreasing at a very slow rate. Hence, I used this as a stopping criterion for gradient descent.

Part-1 EC

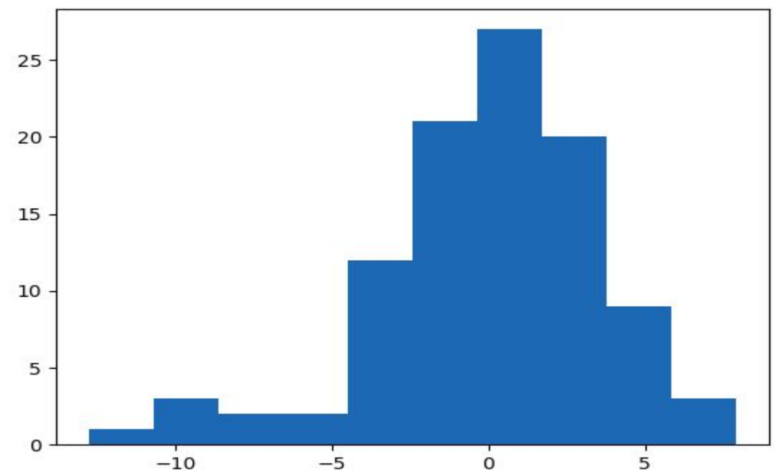
1) Using K- fold validation, the ideal M=9



2) Using Moore penrose inverse matrix-

w= [17.62 -18.27 3.04 61.31 -15.68 -52.51 24.47 9 -6.86   1]
Error- 13.40375753197741

3) Residual error plot-



Upon observing the histogram, we can assume that the error distribution is the linear combination of two gaussian distribution. i.e.

$Z = aX + bY$, $X \sim N(\mu1, \sigma1^2)$, $Y \sim N(\mu2, \sigma2^2)$

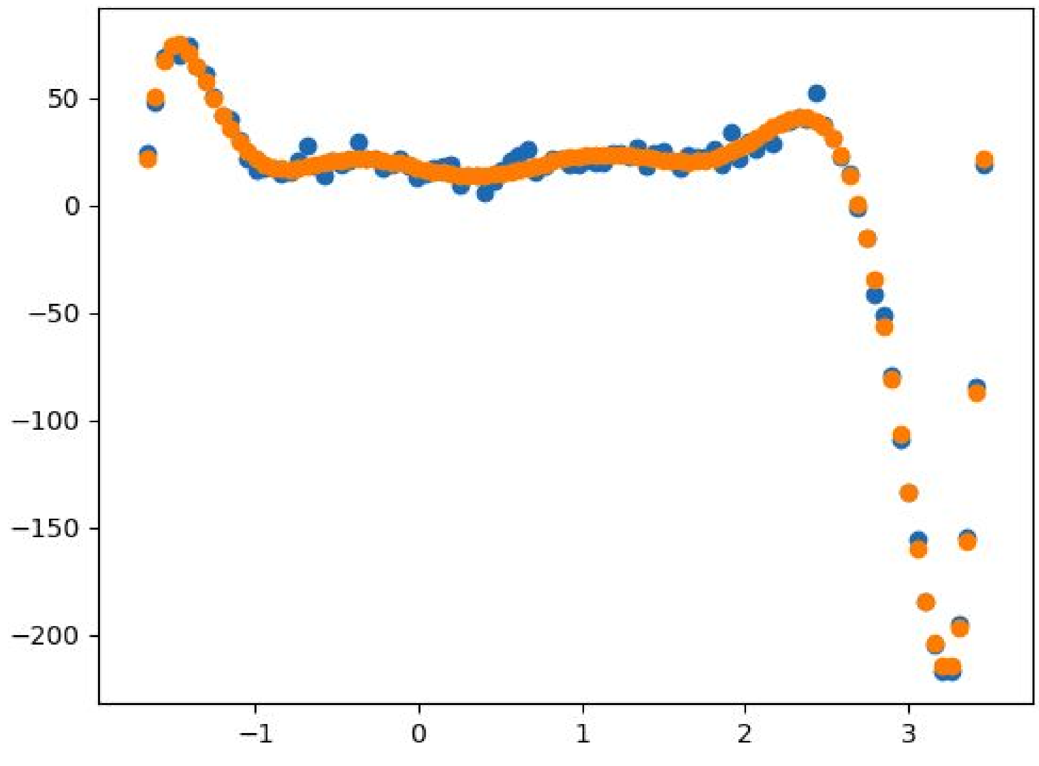
Therefore, $\mu = a\mu1 + b\mu2$

$$\sigma^2 = (a^2)*(\sigma1^2) + (b^2)*(\sigma2^2)$$

From the histogram, it can be observed that $\mu1 = 0.67$, $\mu2 = -9.04$

Also, using the 68–95–99.7 rule, we can say the 68% of the data is between $\mu-\sigma$ to $\mu+\sigma$, hence we get $\sigma1 = 1.77$, $\sigma2 = 0.23$. Also, from the residual errors, we get $\mu=0$, $\sigma^2 = 13.4$. On solving for a, b we get a=2.06, b=0.153.

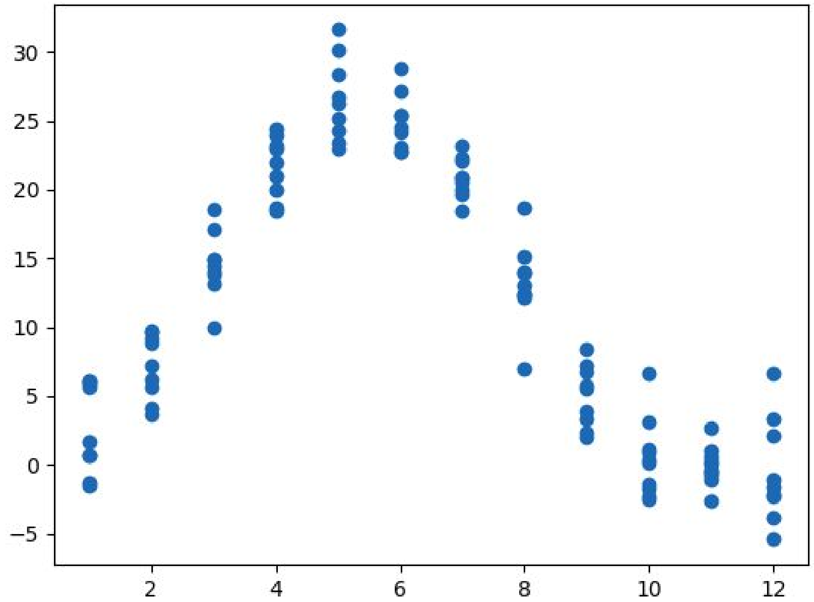
Therefore, $Z = 2.06*X + 0.153*Y$



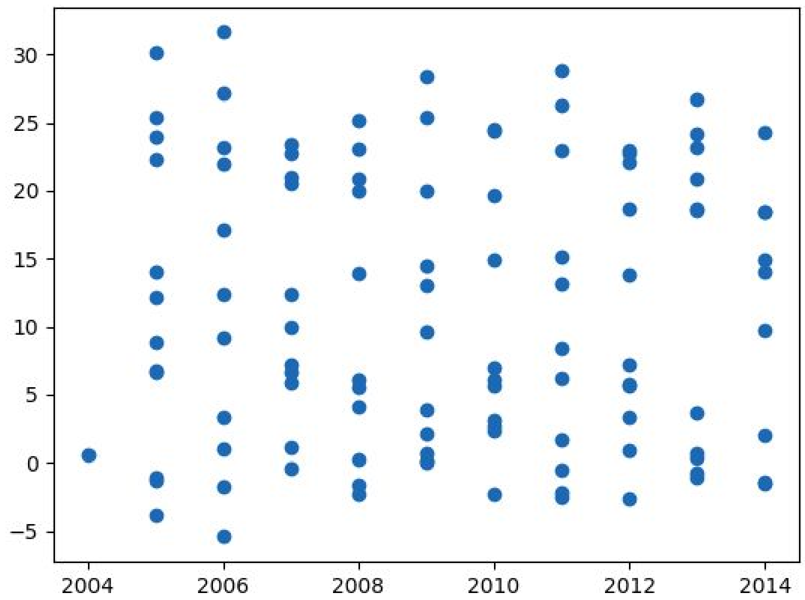
Predicted data vs Actual data

Part 2-

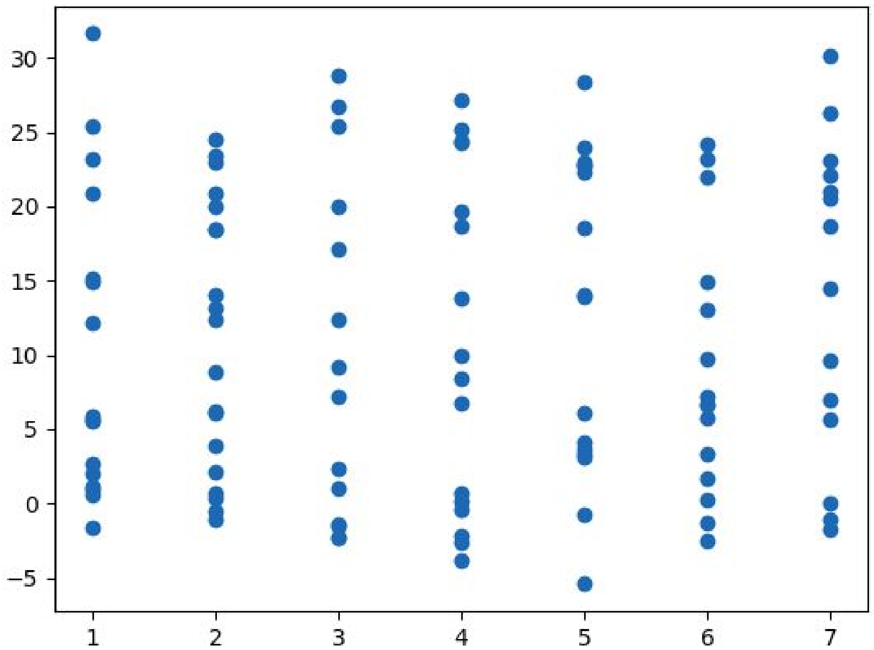
1) Value vs Month-



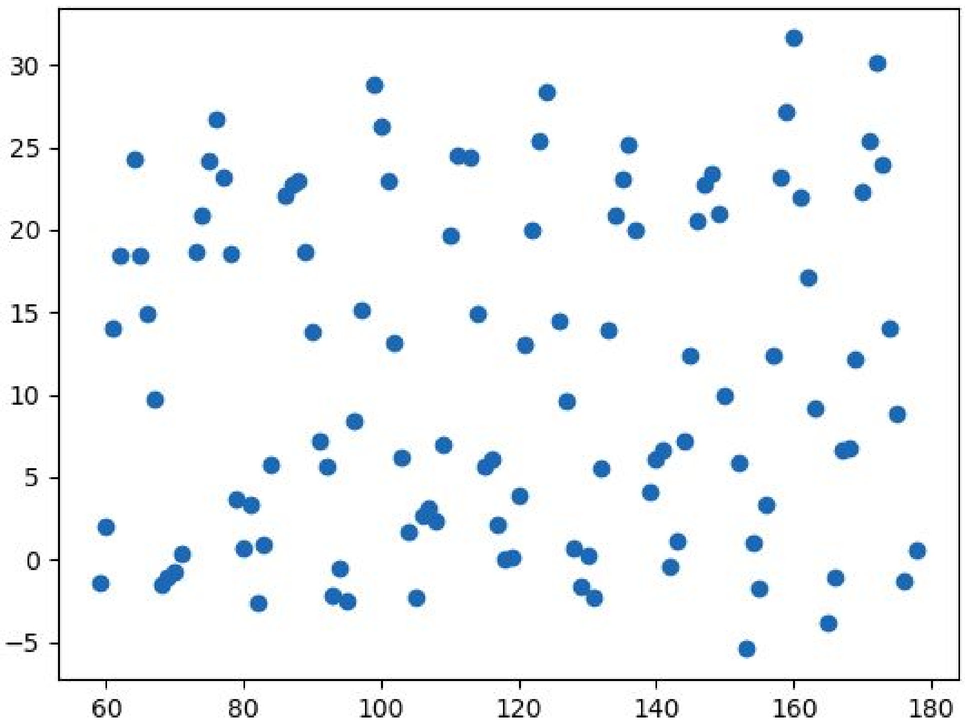
2) Value vs Year-



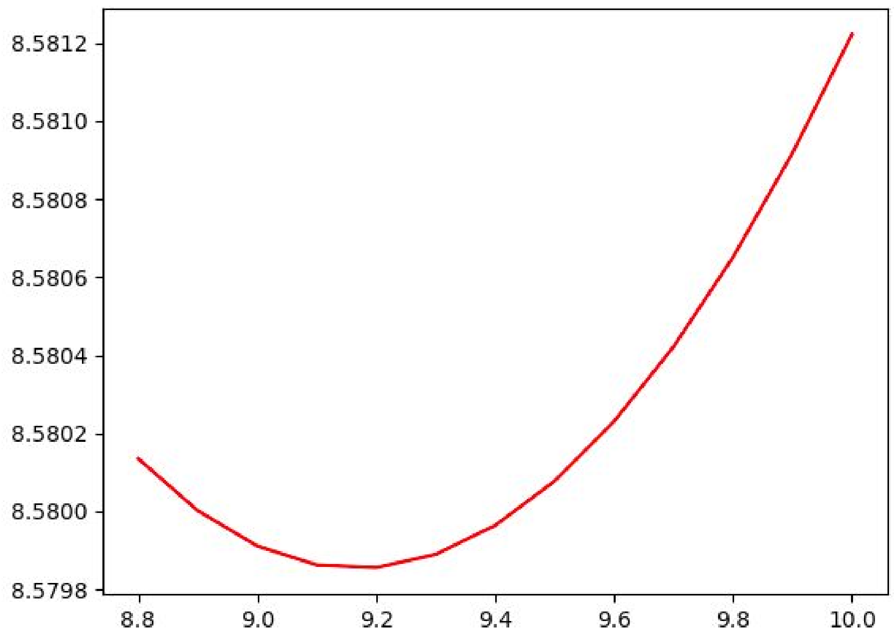
3) Value vs Day

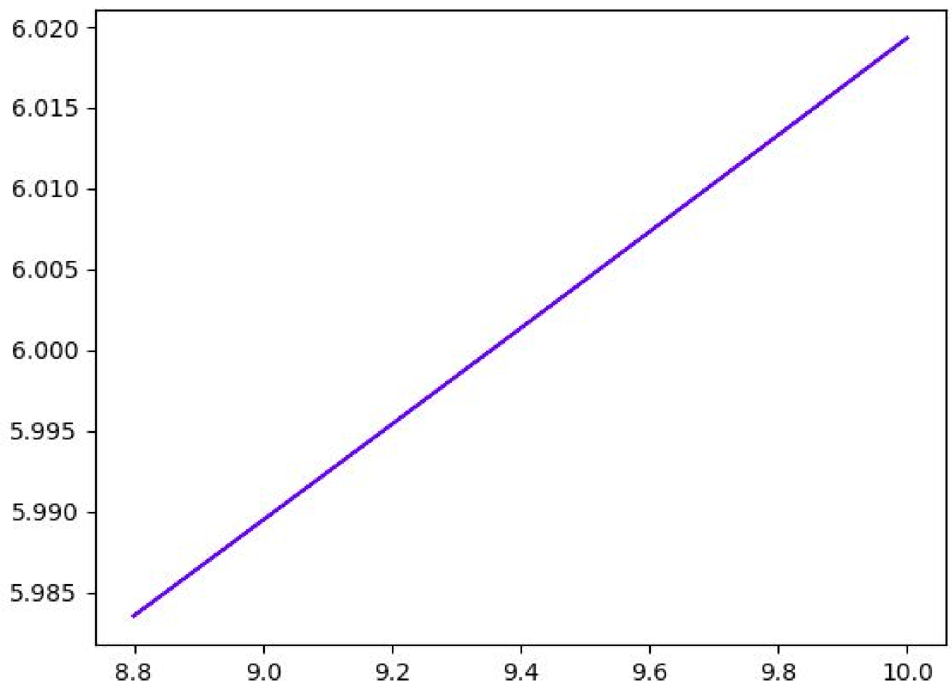


4) Value vs Months Passed-



5) <u>Cross Validation Curves-</u>



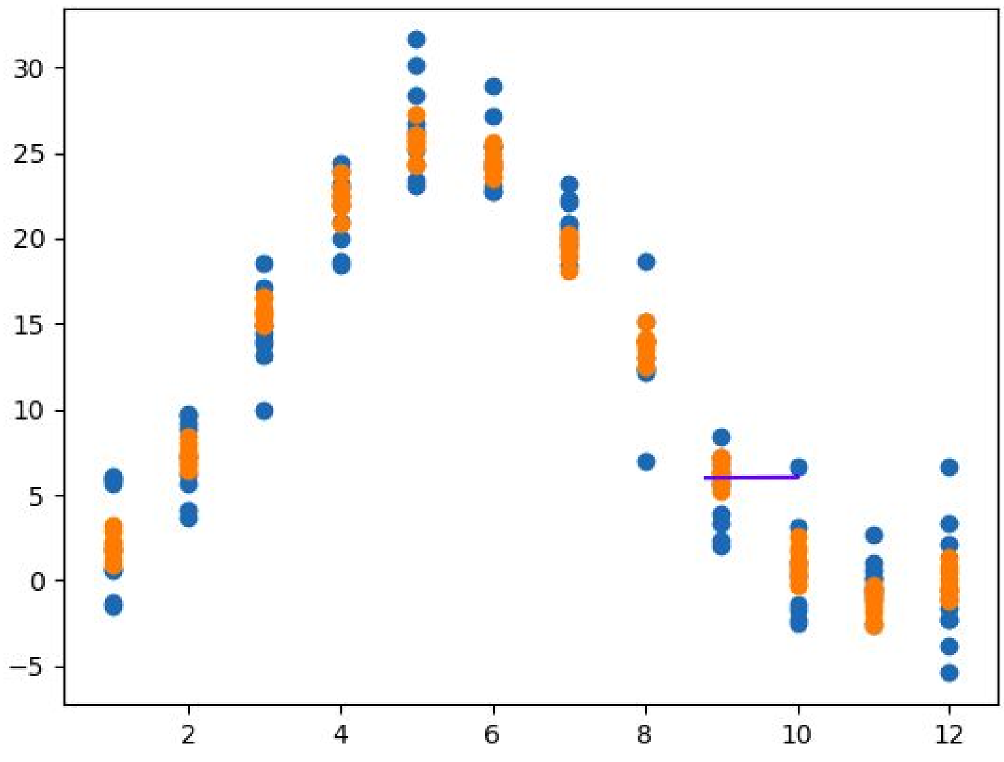


Note: Regularisation constant is around 9.2

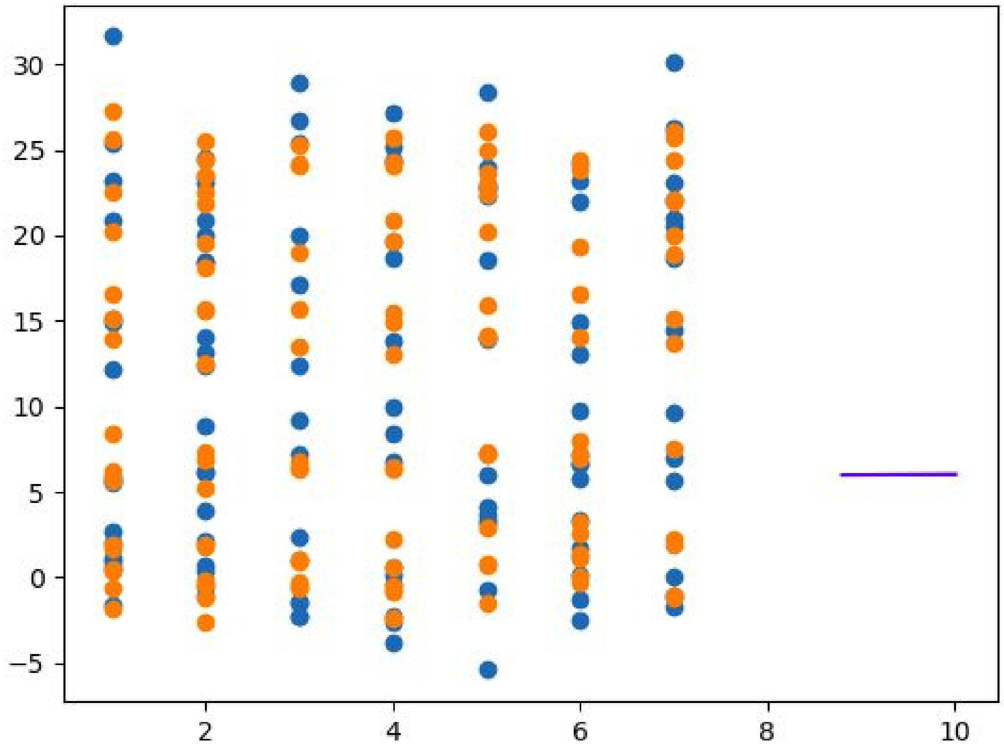
Results-

1) Number of features used- 23
2) w=
   [5.88e-02 4.96e-01 -1.96e-01 3.15e+00 -4.7e-01 8.54e-03
   7.26e-04 4.17e-04 4.19e-03 -5.41e-03 -5.86e-01 1.7e-01
   2.66e-02 -1.44e-02 -1.34e+00 2.02e-01 4.15e-04 4.25e-01
   2.43e-01 -3.97e-01 -4.34e-01 -1.8e-01 -1.8e-01]

Error on training data- 6.260964255354548
Predictions on test data- [25.54 22.04 5.87 1.47 5.88 13.49
22.95 15.34 -0.78 19.69]



Comparative plot of predicted value and actual value vs month



Comparative plot of predicted value and actual value vs day