WINTER OF CODE 5.0

FINAL REPORT

WEEK WISE REPORT:

WEEK 1:

- Began very enthusiastically and completed course 1, week 1 in two days

- Implemented univariate linear regression by the second day and built a Github account.

-Started implementing multi variable linear regression, got stuck at a standstill

-Tried defining cost function and the weights by different methods but failed.

-Dropped the progress because could not find a solution and the labs were not yet unlocked (late financial aid application).

WEEK 2:

-Got involved in family functions and was not able to tend to the project, except progressing in the course a little.

-By end of the week, finished implementing multivariable linear regression and half of polynomial regression.

MID EVALUATION:

-Completed univariate and multivariable linear regression.

-Started implementing Polynomial regression.

WEEK 3:

-Progressed in polynomial regression.

-Implemented polynomial regression but the cost was appearing abnormally low.

-rectified it and completed it by the end of the week

WEEK 4:

-Started reading about logistic regression.

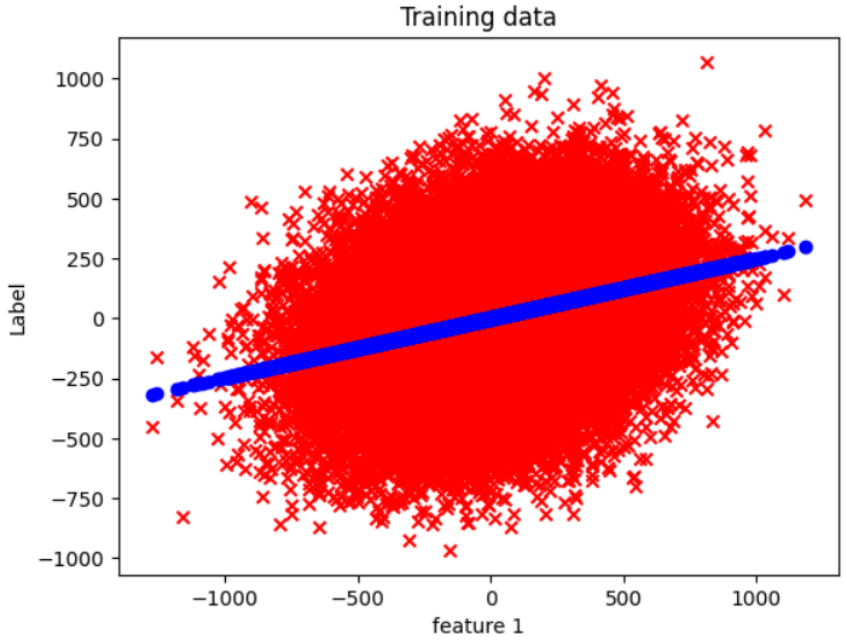
-Started implementation of logistic regression by the end of the week

WEEK 5:

- Dropped the progress for a few days
- Picked up the pace and finished implementation of logistic regression by the end of week.

PROJECT WISE REPORT:

1)LINEAR UNIVARIATE REGRESSION:



LEGENDS:

X=TRAINING DATA

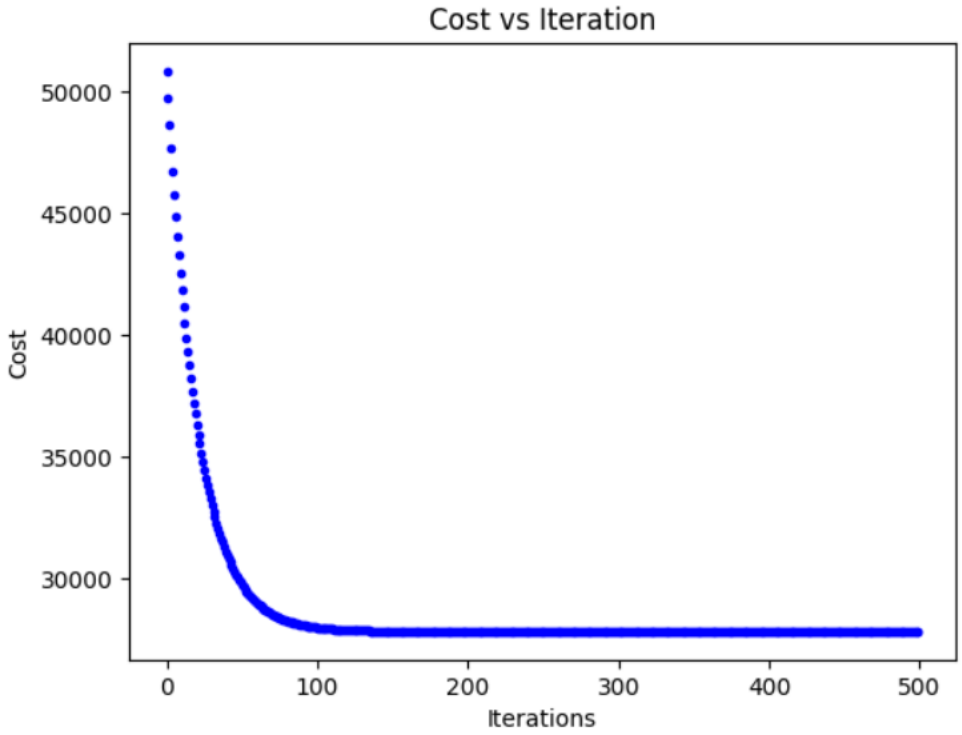
O=PREDICTED DATA

CODE FILE:

LINEAR UNIVARIATE REGRESSION.ipynb

OUTPUT FILE:

linear uni output.csv



Learning rate = 0.0000242

Iterations = 500

-Took a random feature along with the label from the test dataset provided for multivariable linear regression as the training dataset.

- Then trained the model and used it to predict the output for the same dataset used for training the model.

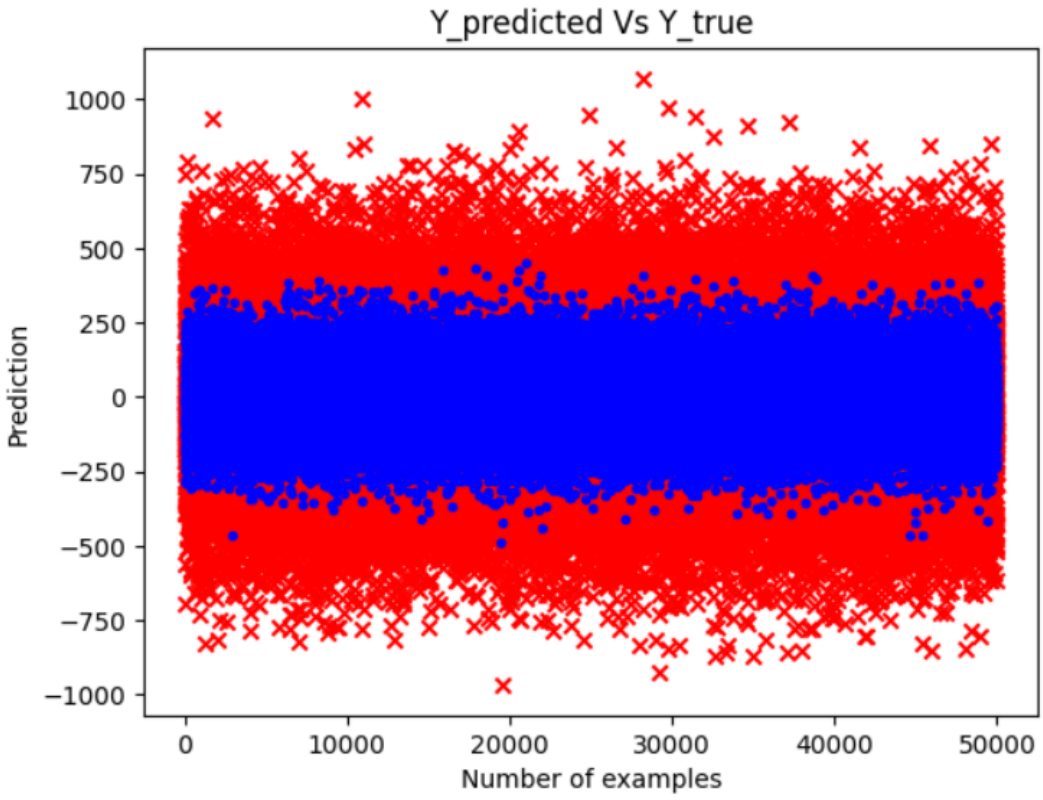
- Was my FIRST VERY OWN MACHINE LEARNING PROJECT!!

2)MULTIVARIABLE LINEAR REGRESSION:

- I implemented this model twice, first without feature scaling and then with feature scaling.

Codefile: Linear_Multivariable_Regression.ipynb; Labelfile: linear outputs.csv

-Without feature scaling:



LEGENDS:

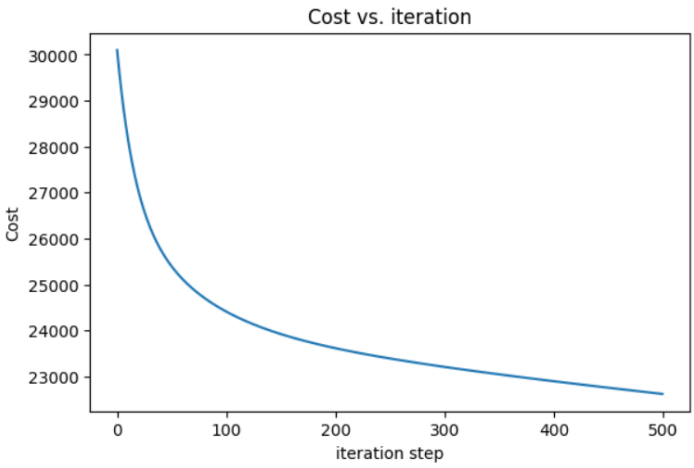
X=TRAINING DATA

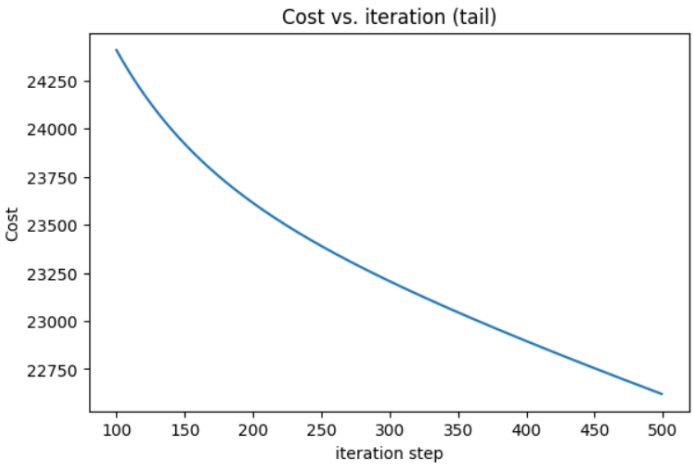
O=PREDICTED DATA

Alpha= 0.0000001

Iterations = 500

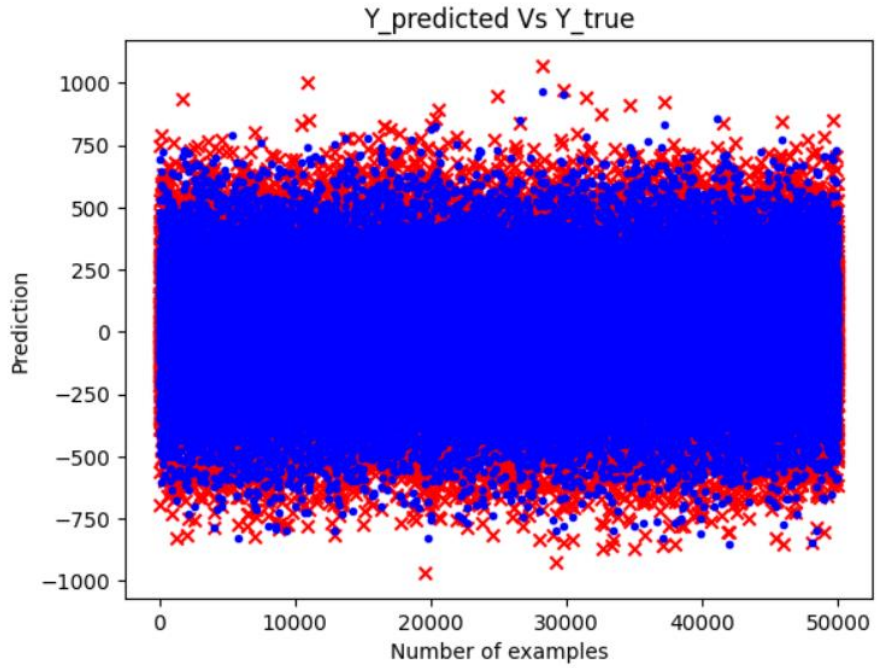
Final Cost ~ 22753.46





- I was kinda stuck at this point and was a bit worried looking at such a high cost value and since it didn't seem to converge I knew something was the problem, Thus I had to use feature scaling.

-WITH FEATURE SCALING:



LEGENDS:

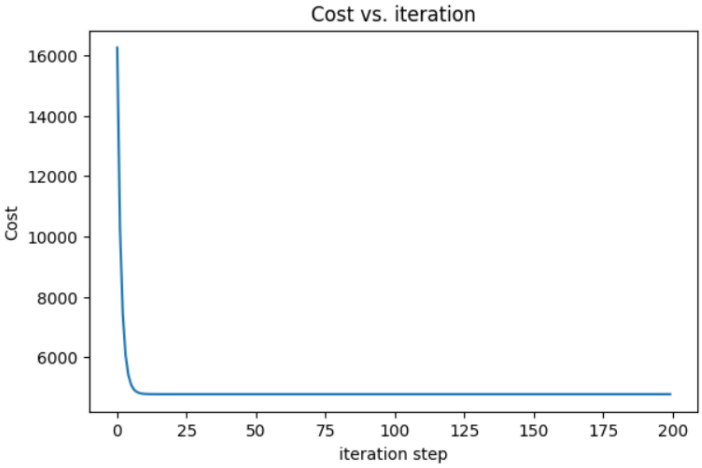
X=TRAINING DATA

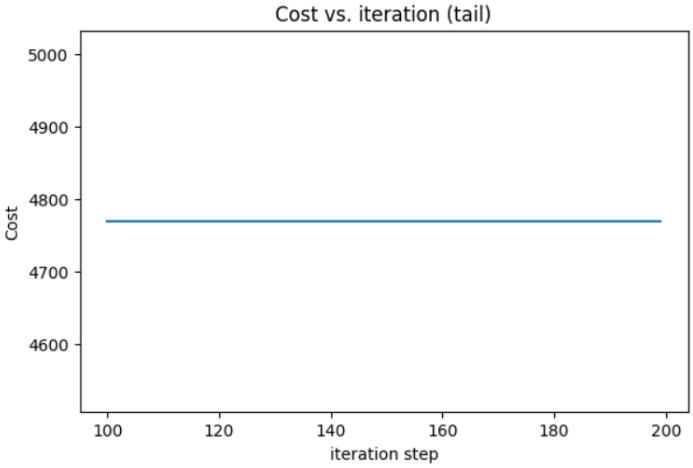
O=PREDICTED DATA

Alpha= 0.0000001

Iterations = 500

Final cost= 4769.77





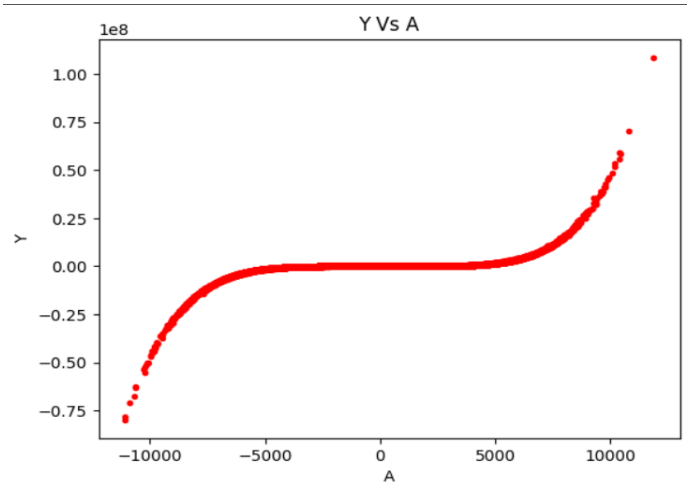
The cost function reduced drastically in the starting iterations, hence the tail has flattened by the end.

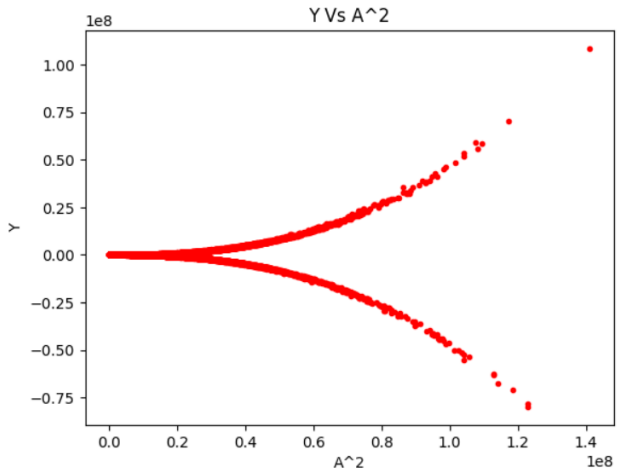
3)POLYNIOMIAL REGRESSION:

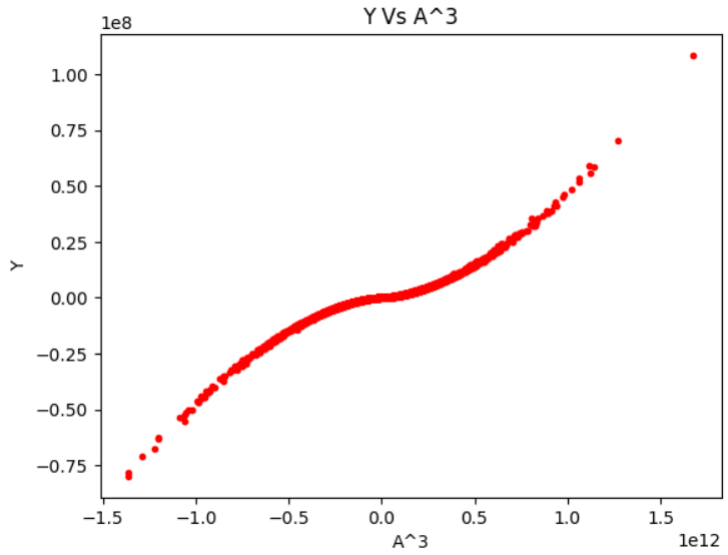
-Codefile: Polynomial Regression.ipynb ; Labelfile: polynomial_output.csv

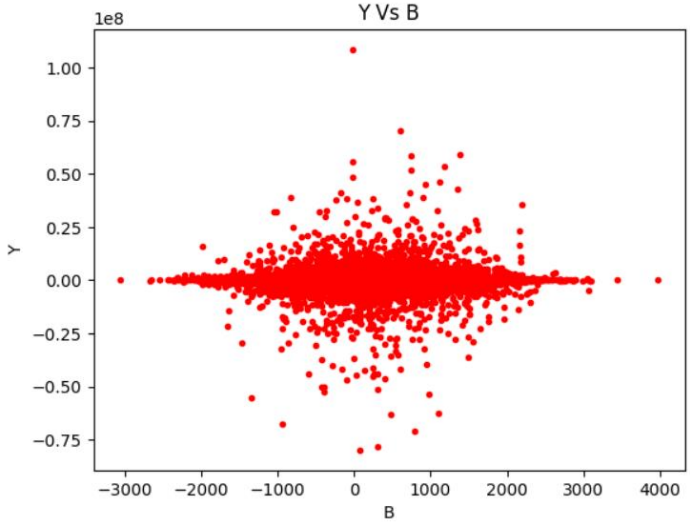
-I chose a degree three polynomial by default and then trained my model based on that.

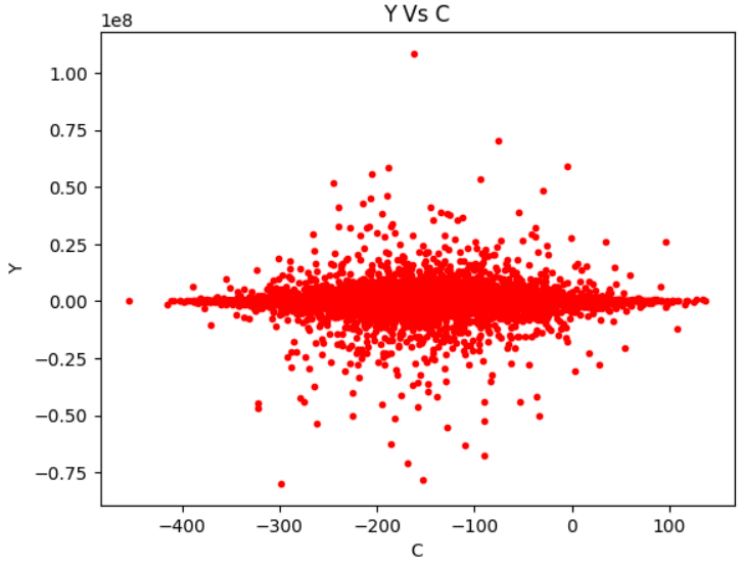
-Visualising data:







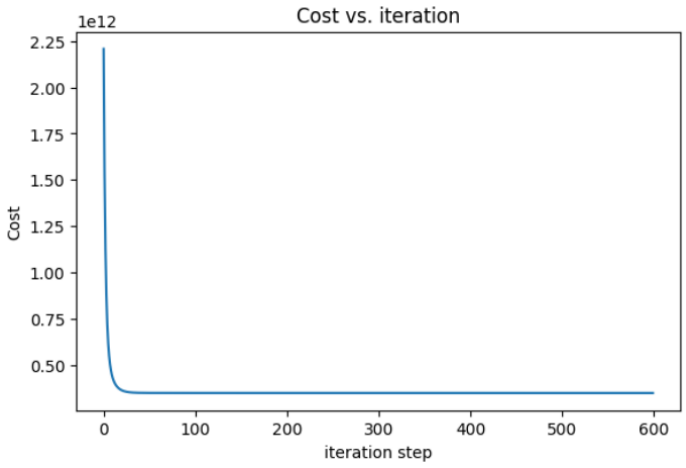


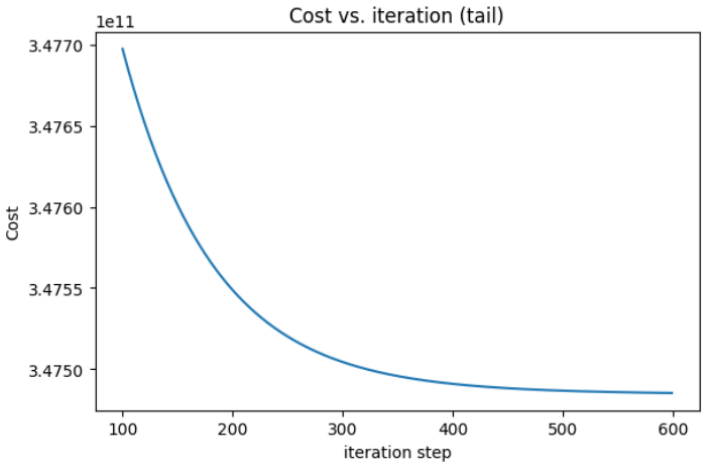


On looking at the pictures, one can say that there seems a relation between the feature A and Y, and that too of cubic nature. Hence we selected our model function to be a cubic polynomial.

Since plots oy labels against B and C do not show any significant results, we will plot the results or predictions against powers of A only.

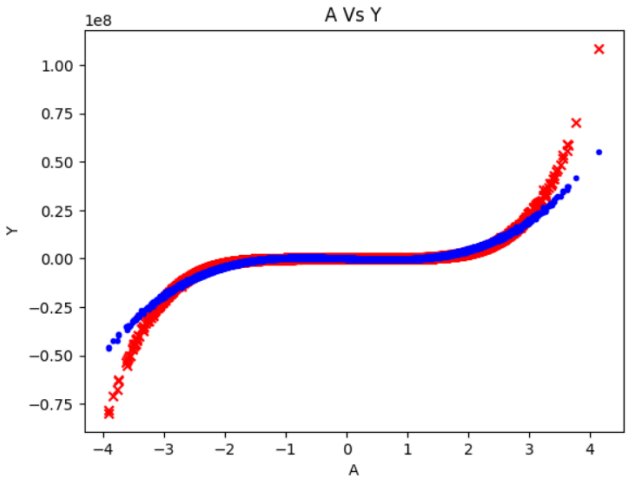
- After visualising the data, choosing the model function and then defining hyperparameters,etc. gradient descent was performed, predictions were made and then cost was calculated.

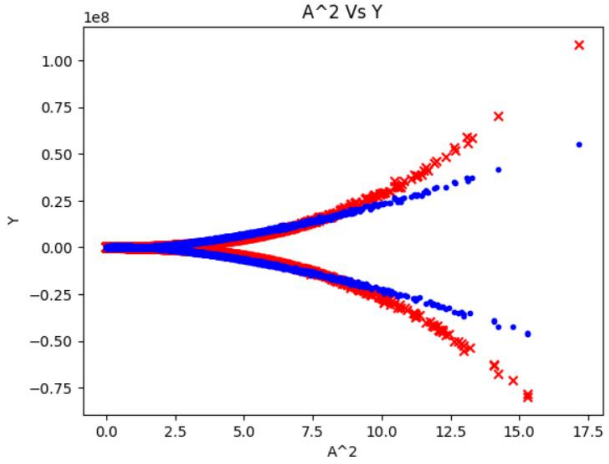


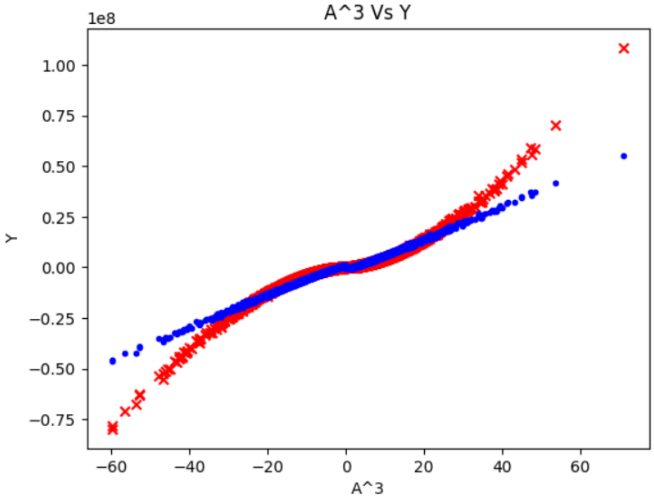


Alpha=0.5 Iterations= 600 Final Cost ~ 347486067289.15 ~ 3*10^12

-Plots of predictions vs true values (Over the training dataset):







LEGENDS:

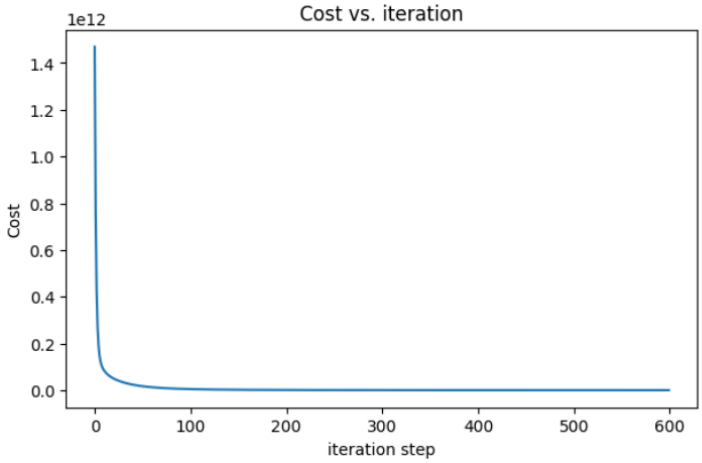
'X' : True Values

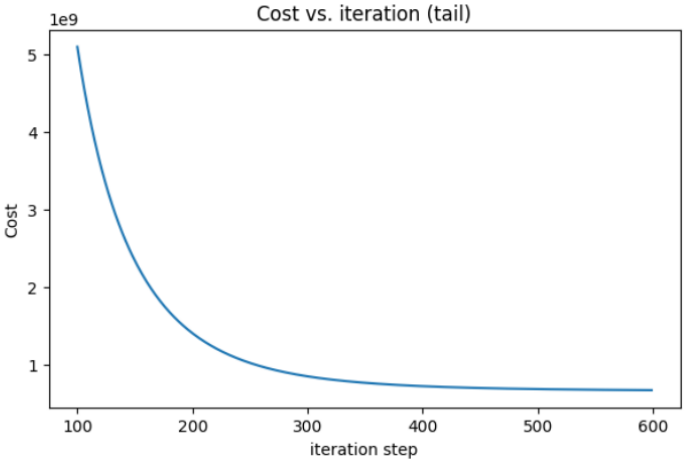
'O' : Predicted Values

R2 score : 0.895878

-Since the predictions still appeared so off, I added some degree 4 and 5 parameters and then re-trained the model, which was quite effective.

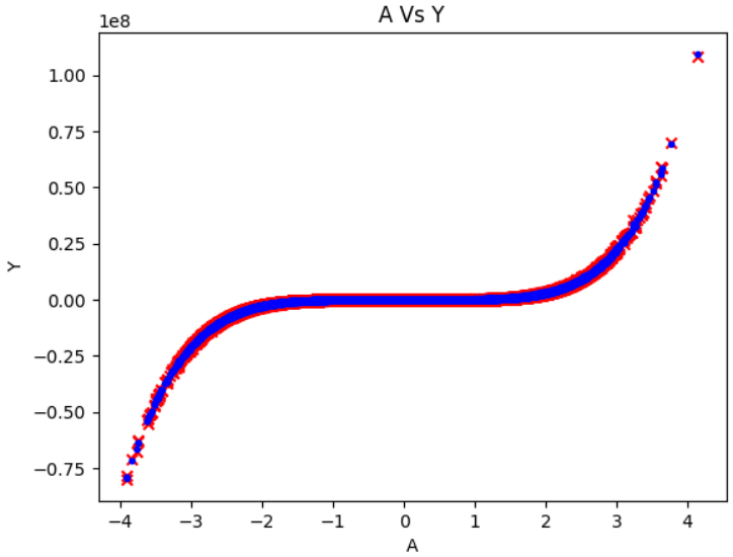
Codefile: Polynomial_Regression deg 5.ipynb;
Labelfile: polynomial out deg 5.csv

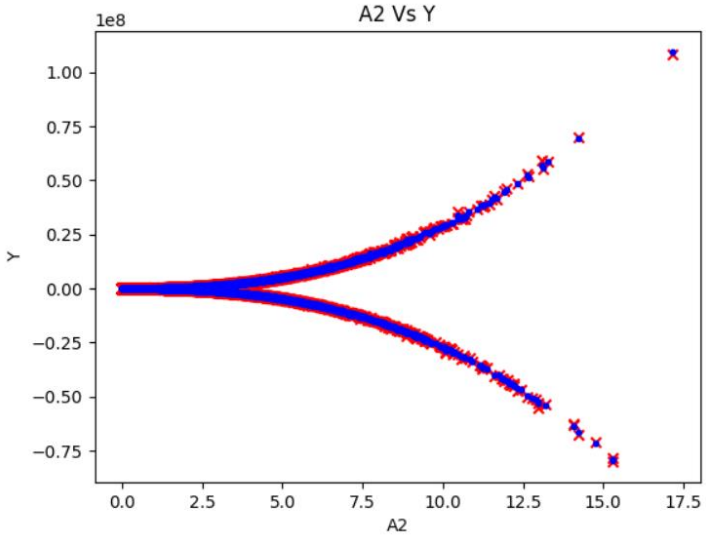


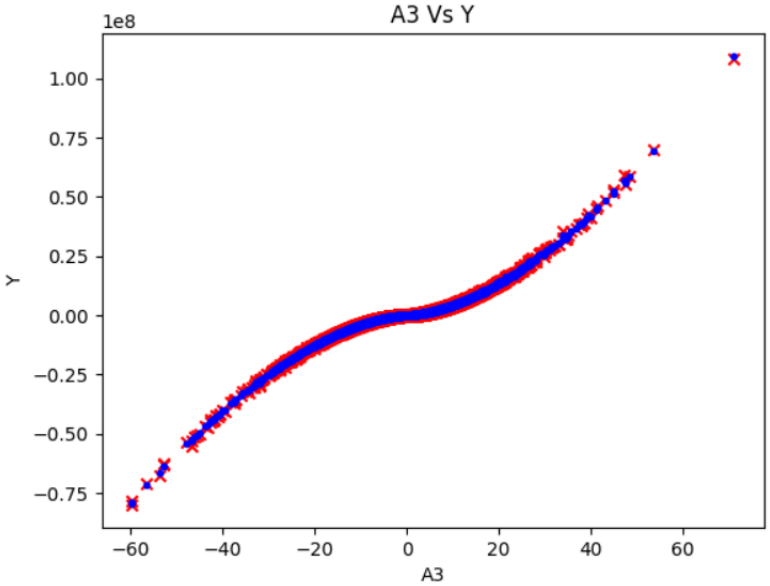


Alpha =0.25 Iterations=600 Final Cost~678624973 ~ 6.7 * 10^8

-PLOTS OF TRUE VALUES VS PREDICTED VALUES:







LEGENDS:

'X' : True Values

'O' : Predicted Values

R2 score : 0.99

THE FEELING OF SATISFACTION
AFTER GETTING SUCH A FIT!!! DAMN

4)LOGISTIC REGRESSION:

- Codefile: Logistic Prediction.ipynb, Logistic regression.ipynb

Labelfile: classification outputs.csv

- Alpha=0.2
- Iterations=50 per binary classification model – repeated for the ten classes.
- Accuracy – 74.46%

----------END OF REPORT----------

Author's Note hehe:

-Why did I choose this project?

-Lately, while I was reading about stock markets; they appeared very intriguing to me and I wanted to know more about the topic; that was when I got to know about quantitative models, predictive models and machine learning itself.

I used to think in awe about how inputting some lines and letters into the machines makes them do these awesome things but didn't bother to do a thorough research by myself; But while researching about the project; when I got a taste of machine learning, It was literally so cool, as if the mist was unshrouding and making programs for machine learning started appearing more humane and possible. That was when I decided that Machine Learning is worth giving a try.

It was a late call though, I was about to go for the Blockchain division, but initially when there was no information of being admitted into it via WOC, I thought to give machine learning a try.

-THE UPS AND DOWNS:

-I started the project very enthusiastically but as always, it faded pretty soon. Once again, I could not finish what I started; these were the thoughts that were looming around but I somehow sticked to it for long enough.

-I remember that after coming back to college, I was a lot demotivated; from the talks and the mockery about ML division being the hardest to get into among all the divisions. At a point of time, I'll admit that at a point in time,I lost the faith and will to be

able to complete the project; but despite all of that, by the end of deadline; I was able to complete Three implementations and The first course; and that was a huge achievement for me, given my past record of these 5 months. Hurraayyyy!

Thus, I would like to thank the CyberLabs team to provide me with this opportunity and my respected mentors and seniors without whom I couldn't have completed even these three projects.

THANK YOU