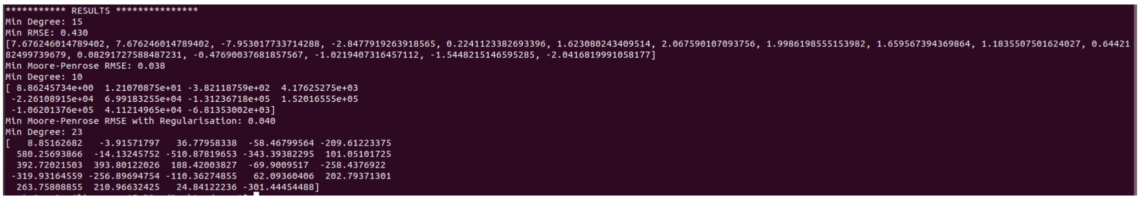
ELL409 Assignment 1

Pratyush Pandey

September 2019

1 Part1 - 20 data points

Using only 20 data points



```
*********** RESULTS ***************
Min Degree: 15
Min RMSE: 0.430
[7.676246014789402, 7.676246014789402, -7.953017733714288, -2.8477919263918565, 0.2241123382693396, 1.623080243409514, 2.067590107093756, 1.9986198555153982, 1.659567394369864, 1.1835507501624027, 0.64421
82499739679, 0.08291727588487231, -0.47690037681857567, -1.0219407316457112, -1.5448215146595285, -2.0416819991058177]
Min Moore-Penrose RMSE: 0.038
Min Degree: 10
[ 8.86245734e+00  1.21070875e+01 -3.82118759e+02  4.17625275e+03
 -2.26108915e+04  6.99183255e+04 -1.31236718e+05  1.52016555e+05
 -1.06201376e+05  4.11214965e+04 -6.81353002e+03]
Min Moore-Penrose RMSE with Regularisation: 0.040
Min Degree: 23
[   8.85162682   -3.91571797   36.77958338  -58.46799564 -209.61223375
  580.25693866  -14.13245752 -510.87819653 -343.39382295  101.05101725
  392.72021503  393.80122026  188.42003827  -69.9009517  -258.4376922
 -319.93164559 -256.89694754 -110.36274855   62.09360406  202.79371301
  263.75808855  210.96632425   24.84122236 -301.44454488]
```

Figure 1: The estimated polynomial coefficients for all the methods - along with the RMS errors. The Moore-Penrose pseudoinverse gives the best result, with the regularised version following closely behind

1.1 Without Regularisation

1.1.1 Using Moore-Penrose Pseudoinverse

Cross Validation is used to approximate the testing error. The graph between cross validation error and training error versus degree is shown as below. The random seed is set as 1 for all the subsequent observatons.

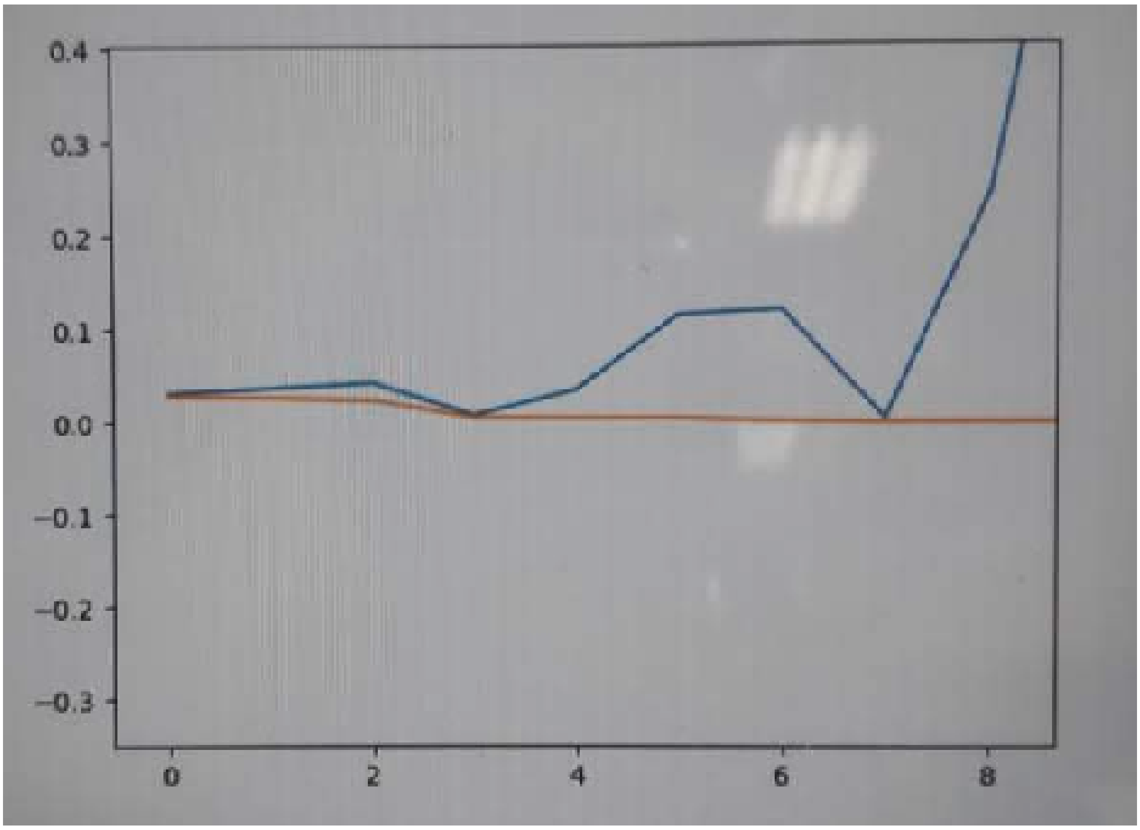


Figure 2: The blue line is cross validation error, the orange line is training error. The sweet is found as degree=10

The polynomial found for this sweet spot is

$$y = 8.86 + 12.1x - 382.11x^2 + 4176.25x^3 - 22610.89x^4 - 69918.3225x^5 - 131236.718x^6 +$$

$$152016.5x^7 - 106201.37x^8 + 41121.49x^9 - 6813.53x^10$$

The coefficient matrix is - [8.86245734e+00 1.21070875e+01 -3.82118759e+02 4.17625275e+03 -2.26108915e+04 6.99183255e+04 -1.31236718e+05 1.52016555e+05 -1.06201376e+05 4.11214965e+04 -6.81353002e+03]

The noise variance is 0.001444
The plot of this polynomial along with the scatter plot of the points is given as -

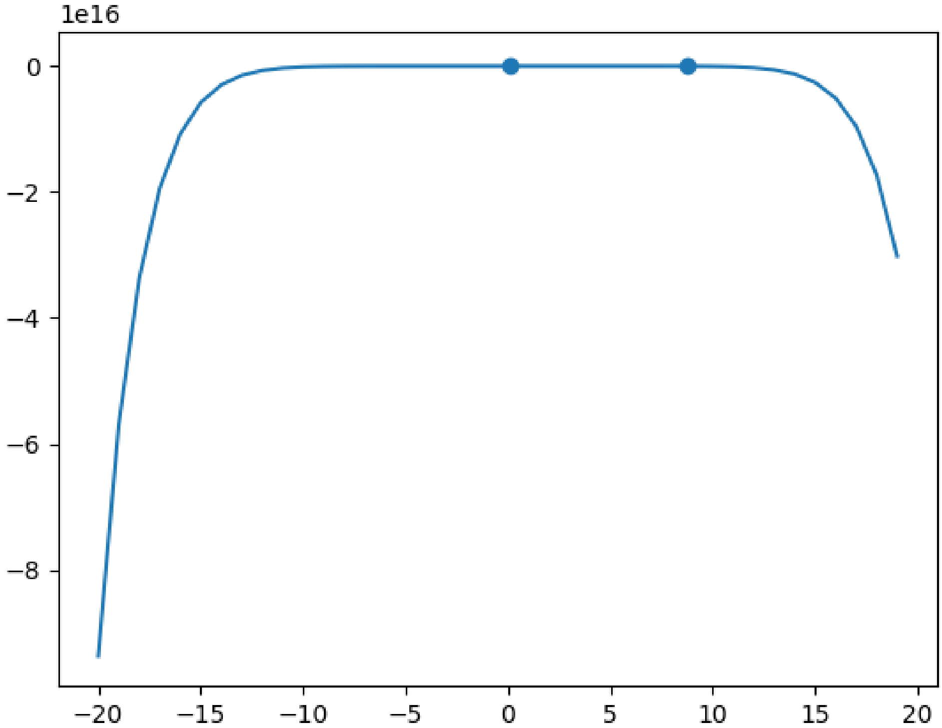
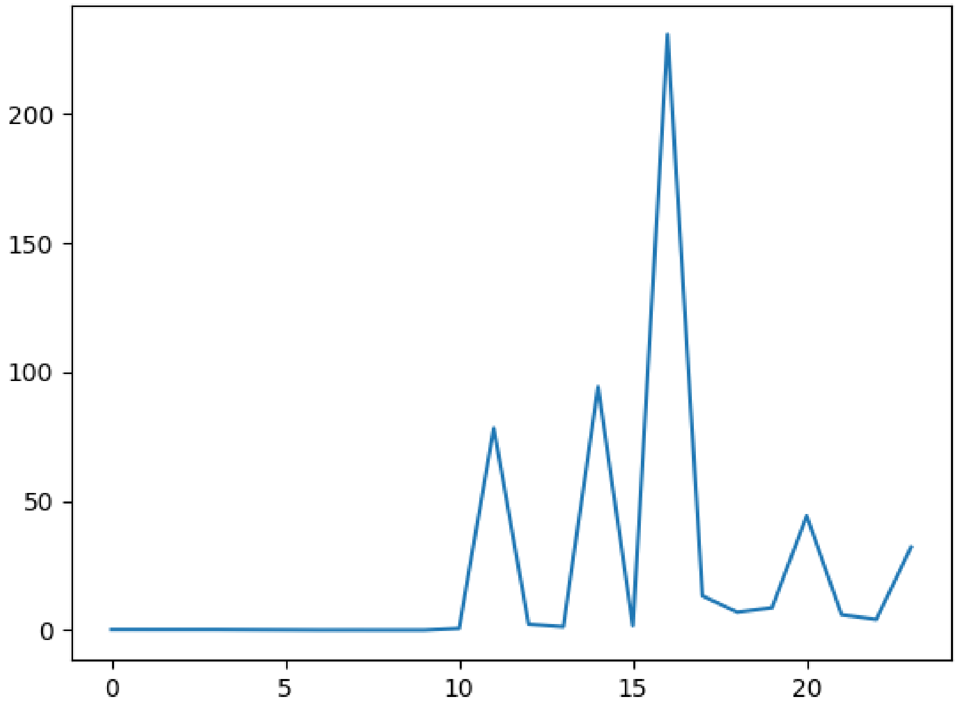


Figure 3: The points are closely located and the graph overfits, at degree 10 for 20 points

This curve seems to show overfitting for x=0 and 10. The degree 4 polynomial may also be a good fit as Cross validation error is also very low for it.



1.1.2 Using Gradient Descent

The Hyperparameters used to train is shown as below

- Learning Rate = Fixed at 0.01
- Epochs = 50
- Stopping Criterion = 5000 iterations
- Batch Size = 1 (SGD), variable (¡=N, the total number of training example in a single fold)
- No of folds for n-folds cross validation = 5

Data Preprocessing was done using Min-Max Normalisation. A random seed of 2 was used. Variation of Cross Validation error with degree of the polynomial used is shown as below.

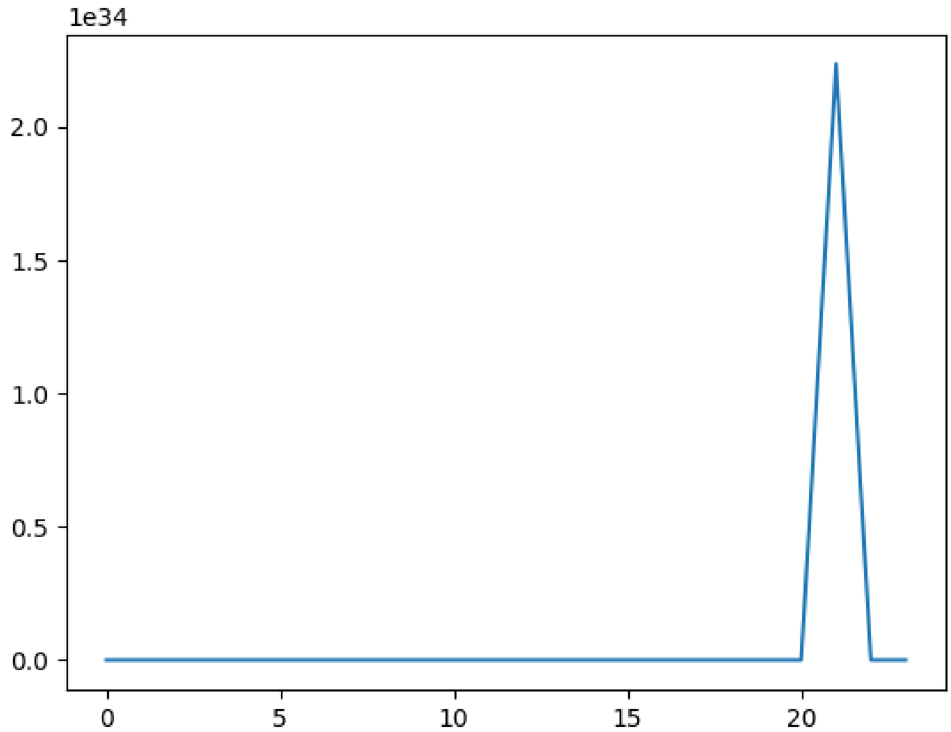


Figure 4: The following graph shows the variation in converging errors with degree of the polynomial. The blue line is cross validation error, the orange line is training error. The sweet is found as degree=15, but isnt visible due to the scaling

The coefficients were - [8.86245734e+00 1.21070875e+01 -3.82118759e+02 4.17625275e+03 -2.26108915e+04 6.99183255e+04 -1.31236718e+05 1.52016555e+05 -1.06201376e+05 4.11214965e+04 -6.81353002e+03]

The variance of the underlying noise was estimated to be 0.053361.

The following graph plots the predicted fifteenth degree polynomial with the points in consideration.

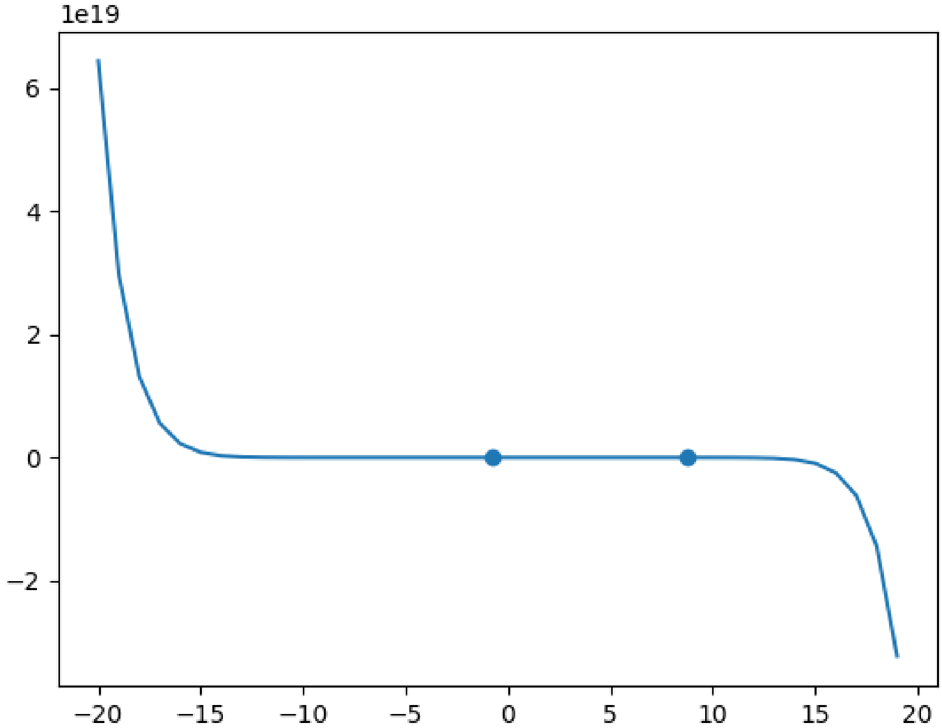
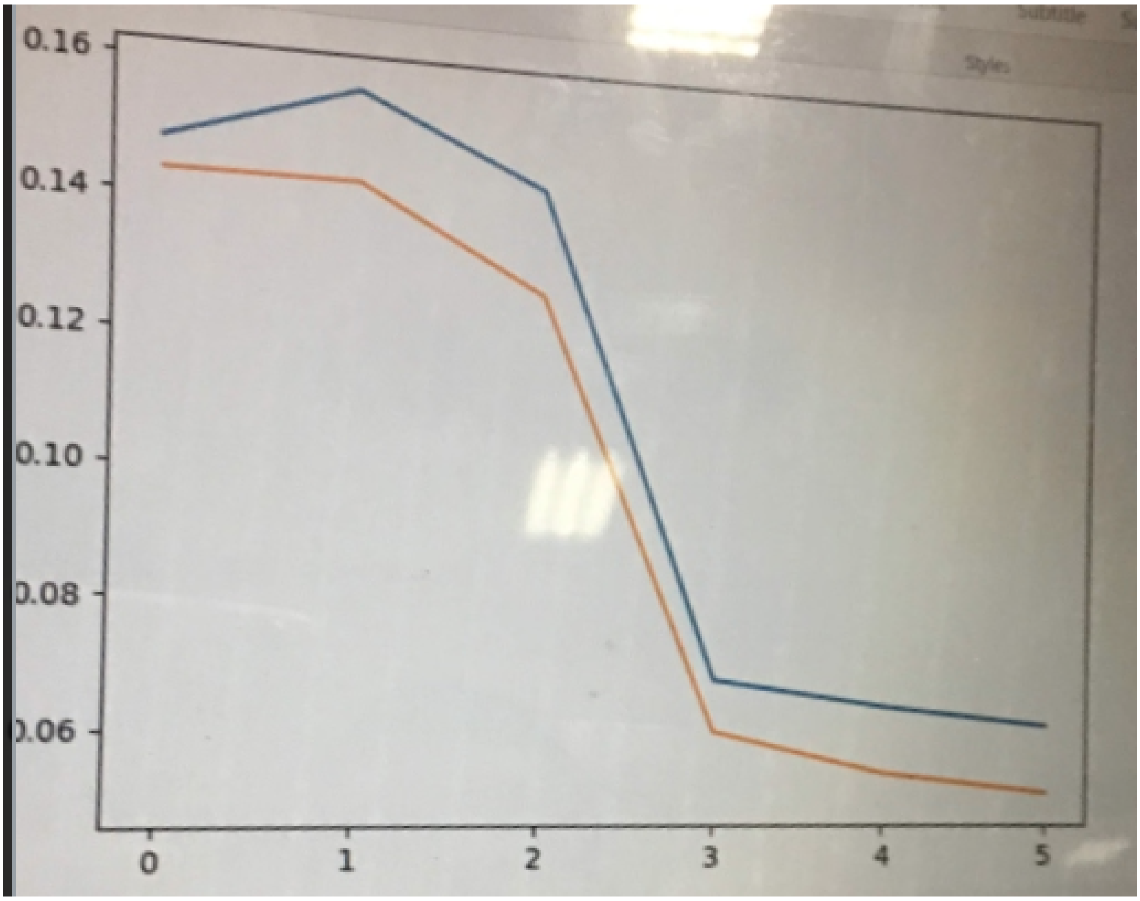


Figure 5: The blue line is cross validation error, the orange line is training error. The sweet is found as degree=15

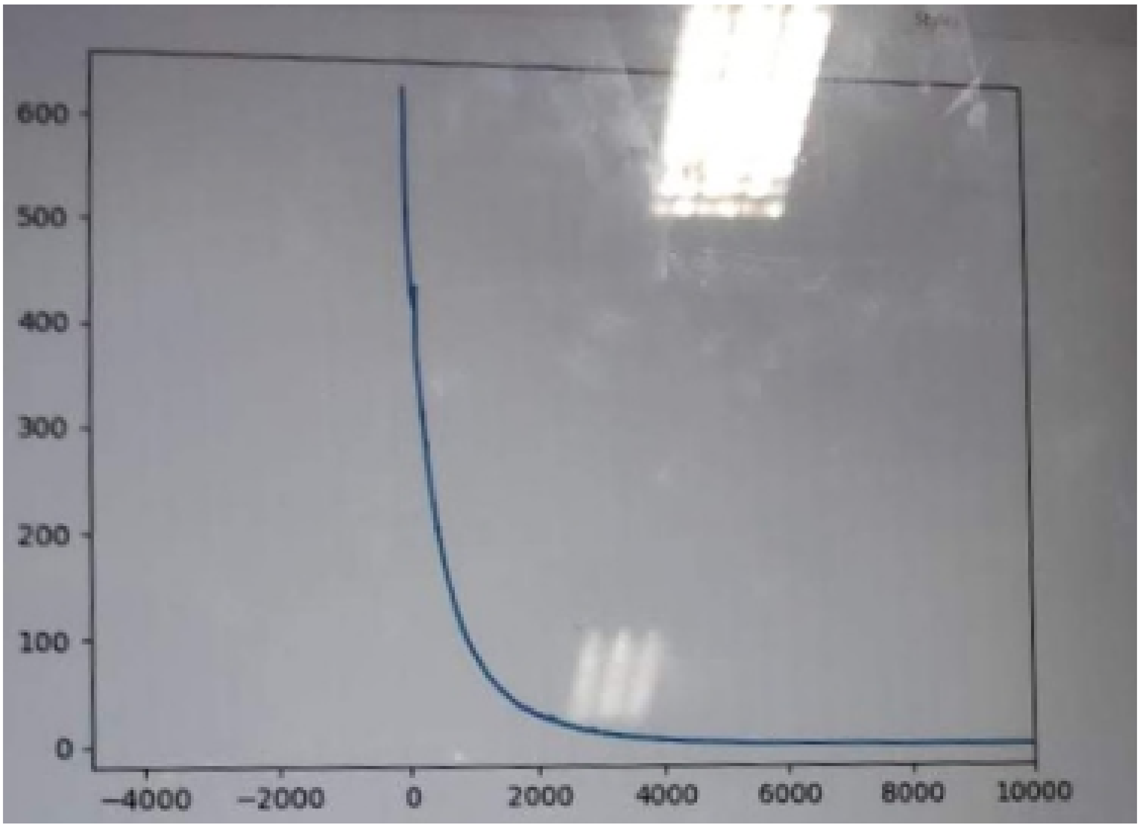
This seems to be a good fit but is actually a little underfit as higher degree estimated were not obtained.

The gradient stopped converging after degree 20 while cross validation is still decreasing, so the obtained result that will be not good.

The following graph shows convergence of the Gradient descent algorithm for degree 5.



For degree 15, graph converges quicker than when the batch size was kept at 1.



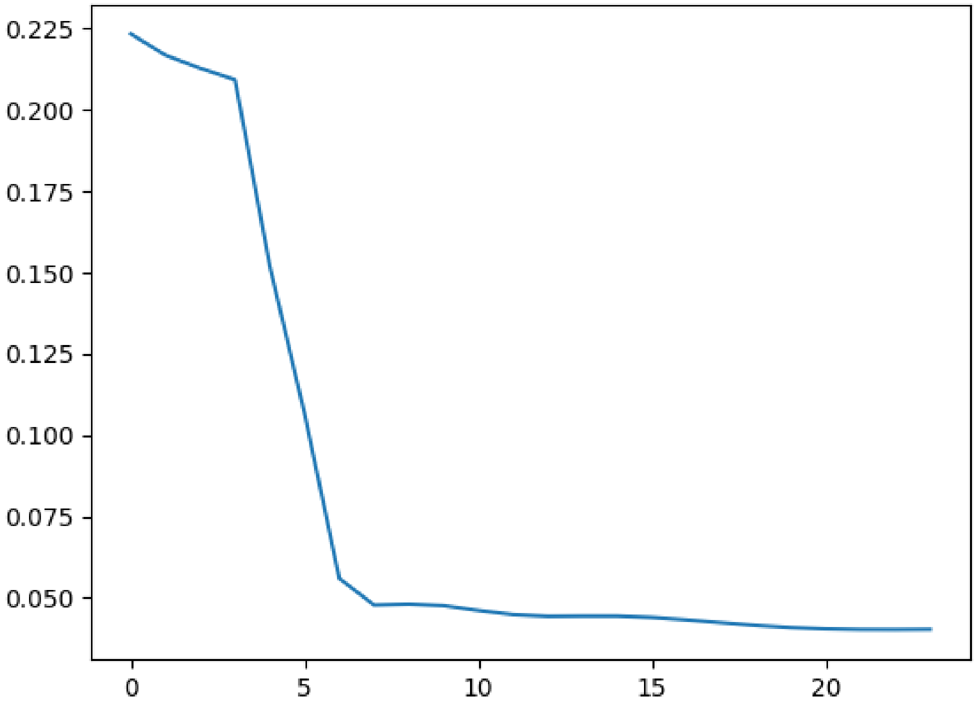
1.2 With Regularisation

1.2.1 Moore-Penrose Pseudoinverse

The cross validation and training error vs ln(lambda) is as shown below.

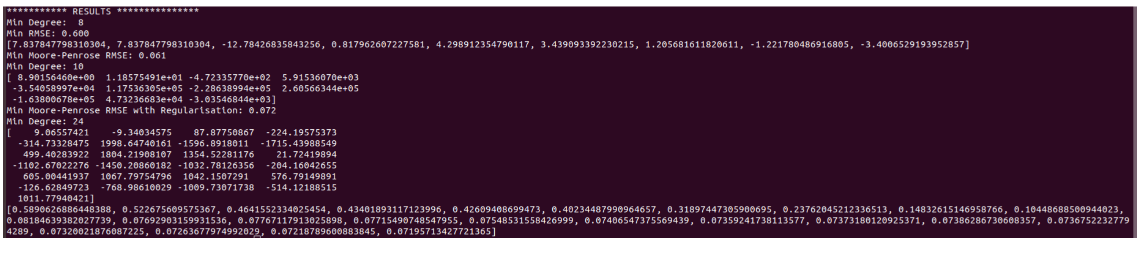
The least cross error is obtained at lambda = 1e-7 Corresponding degree of polynomial = 23 The coefficients are shown below - [8.85162682 -3.91571797 36.77958338 -58.46799564 -209.61223375 580.25693866 -14.13245752 -510.87819653 -343.39382295 101.05101725 392.72021503 393.80122026 188.42003827 -69.9009517 -258.4376922 -319.93164559 -256.89694754 -110.36274855 62.09360406 202.79371301 263.75808855 210.96632425 24.84122236 -301.44454488]

Noise Variance = 0.00016



2 Part 2 - 100 data points

2.1 Without Regularisation



```
*********** RESULTS ***************
Min Degree:  8
Min RMSE: 0.600
[7.837847798310304, 7.837847798310304, -12.78426835843256, 0.817962607227581, 4.298912354790117, 3.439093392230215, 1.205681611820611, -1.221780486916805, -3.4006529193952857]
Min Moore-Penrose RMSE: 0.061
Min Degree: 10
[ 8.90156460e+00  1.18575491e+01 -4.72335770e+02  5.91536070e+03
 -3.54058997e+04  1.17536305e+05 -2.28638994e+05  2.60566344e+05
 -1.63800678e+05  4.73236683e+04 -3.03546844e+03]
Min Moore-Penrose RMSE with Regularisation: 0.072
Min Degree: 24
[    9.06557421    -9.34034575    87.87750867  -224.19575373
  -314.73328475  1998.64740161 -1596.8918011  -1715.43988549
   499.40283922  1804.21908107  1354.52281176    21.72419894
 -1102.67022276 -1450.20860182 -1032.78126356  -204.16042655
   605.00441937  1067.79754796  1042.1507291    576.79149891
  -126.62849723  -768.98610029 -1009.73071738  -514.12188515
  1011.77940421]
[0.5890626886448388, 0.522675609575367, 0.4641552334025454, 0.43401893117123996, 0.42609408699473, 0.40234487990964657, 0.31897447305900695, 0.23762045212336513, 0.14832615146958766, 0.10448688500944023,
0.08184639382027739, 0.07692903159931536, 0.07767117913025898, 0.07715490748547955, 0.07548531558426999, 0.07406547375569439, 0.07359241738113577, 0.07373180120925371, 0.07386286730608357, 0.0736752232779
4289, 0.07320021876087225, 0.07263677974992029, 0.07218789600883845, 0.07195713427721365]
```

2.1.1 Using Moore-Penrose Pseudoinverse

Cross Validation is used to approximate the testing error. The graph between cross validation error and training error versus degree is shown as below. The random seed is set as 1 for all the subsequent observatons. Min Moore-Penrose RMSE: 0.061 Min Degree: 10 [8.90156460e+00

1.18575491e+01 -4.72335770e+02 5.91536070e+03 -3.54058997e+04 1.17536305e+05 -2.28638994e+05 2.60566344e+05 -1.63800678e+05 4.73236683e+04 -3.03546844e+03]

2.1.2 Gradient Descent

The Hyperparameters used to train is shown as below

- Learning Rate = Fixed at 0.01
- Epochs = 50
- Stopping Criterion = 5000 iterations
- Batch Size = 1 (SGD)
- No of folds for n-folds cross validation = 5

Min Degree: 8 Min RMSE: 0.600 [7.837847798310304, 7.837847798310304, -12.78426835843256, 0.817962607227581, 4.298912354790117, 3.439093392230215, 1.205681611820611, -1.221780486916805, -3.4006529193952857]

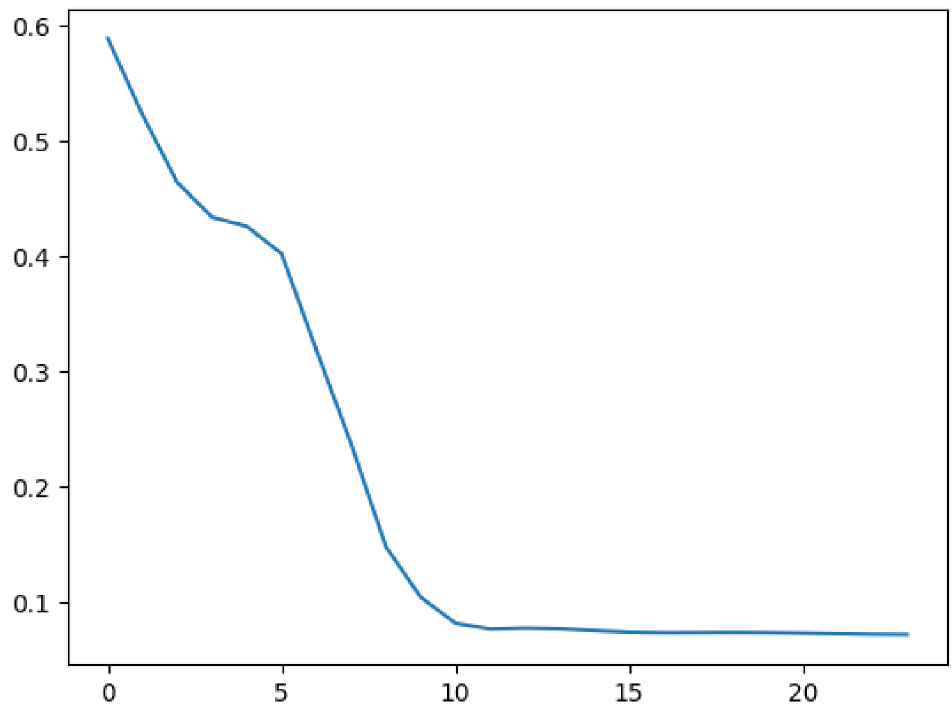
2.2 With Regularisation

2.2.1 Moore-Penrose Pseudoinverse

Min Moore-Penrose RMSE with Regularisation: 0.072 Min Degree: 24 [9.06557421 -9.34034575 87.87750867 -224.19575373 -314.73328475 1998.64740161 -1596.8918011 -1715.43988549 499.40283922 1804.21908107 1354.52281176 21.72419894 -1102.67022276 -1450.20860182 -1032.78126356 -204.16042655 605.00441937 1067.79754796 1042.1507291 576.79149891 -126.62849723 -768.98610029 -1009.73071738 -514.12188515 1011.77940421]

The cross validation and training error vs ln(lambda) is as shown below.

-



rmse$_regu_mp.png$

3 Conclusion

3.1 Difference between the n=20 and n=100 results

In both the cases, similar polynomials were obtained finally (both degree 7 and coefficients were also reasonably close). Hence the final results were approximately the same.

The main difference were obtained in gradient descent. The iterations required in the case of batch size 1 (SGD) were lower for n=20 than the iterations required for n=100. But considering full batch gradient descent, the n=100 case required drastically lesser number of iterations than n=20. This may be because degree of the polynomial in case of n=20 was 7 while it was 5 in case of n=100. Also inclusion of more points might have given more directivity to the gradient making the descent faster (batch size 100 versus batch size 20).

The noise variance obtained in the case of n=100 was slightly greater than n=10. This maybe because more points will mean more noise to capture.

3.2 Final Estimate of the underlying polynomial

The final estimate is chosen (of course in the case of n=100)n from the results of . Following reasons justify this choice -

- The degree for regularisation is very high compared to degree of non regularisation.
- The cross validation error is lesser in the case of non regularisation.

Hence, the final estimate is polynomial has Min Moore-Penrose RMSE: 0.061 Min Degree: 10 [8.90156460e+00 1.18575491e+01 -4.72335770e+02 5.91536070e+03 -3.54058997e+04 1.17536305e+05 -2.28638994e+05 2.60566344e+05 -1.63800678e+05 4.73236683e+04 -3.03546844e+03]

The final noise variance is (estimated) - 0.003721