REPORT FOR ASSIGNMENT 1

TASK1:

- **Brief on what the LinearRegression().fit() will do**

  LinearRegression().fit() tries to best Fit for the given set of features in X and an output Set in y
  Here arguments are in the form

- X = [m different instances of [f1,f2,f3,f4 ………,fn]] ,Here X contains “m” quantities of “n” features.
- Y = [[y1],[y2],[y3],[y4] ……………….[ym]], Y should contain OUTPUT for “m” quantities in X here output contains only one feature that is output
- fit_intercept: can be used to tell the model whether to predict the coefficients with taking intercept into account or not

.fit() ultimately tries to solve the equation

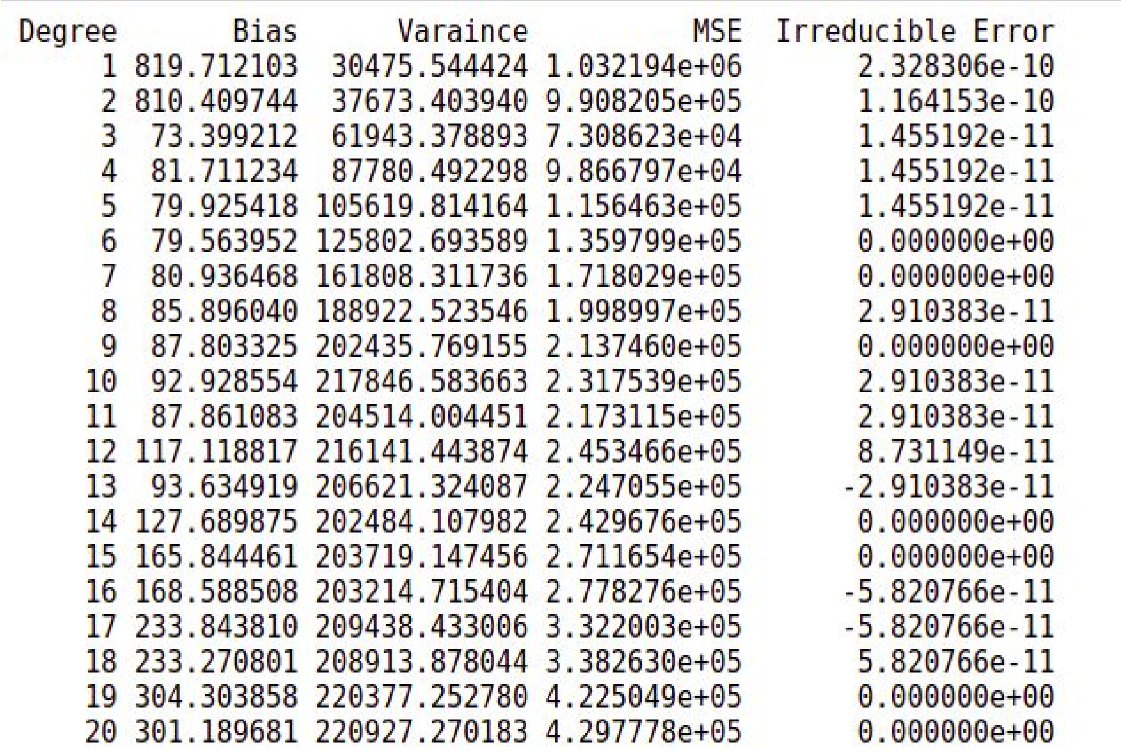
y= a0.x0+a1.x1+a2.x2+a3.x3+.......an.x^n
Where x0=x^0, x1=x^1 … are features of the data model

which can predict the test data and can extract ai’s

Here we have 800 equations for each dataset. And we have n+1 coefficients to find for each polynomial of degree n. This function tries to find the best possible coefficients to fit the 800 equations.

TASK2:

- ## Bias and variance

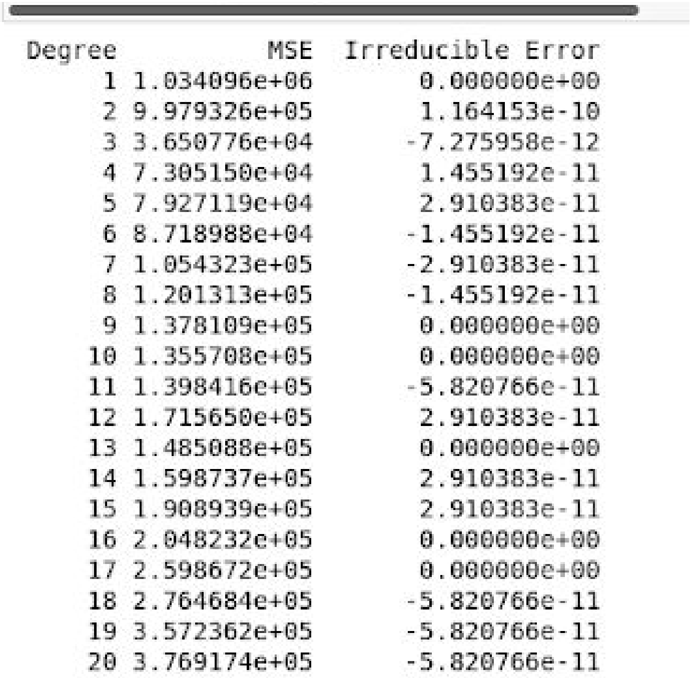


```
Degree        Bias       Varaince          MSE  Irreducible Error
     1 819.712103  30475.544424 1.032194e+06       2.328306e-10
     2 810.409744  37673.403940 9.908205e+05       1.164153e-10
     3  73.399212  61943.378893 7.308623e+04       1.455192e-11
     4  81.711234  87780.492298 9.866797e+04       1.455192e-11
     5  79.925418 105619.814164 1.156463e+05       1.455192e-11
     6  79.563952 125802.693589 1.359799e+05       0.000000e+00
     7  80.936468 161808.311736 1.718029e+05       0.000000e+00
     8  85.896040 188922.523546 1.998997e+05       2.910383e-11
     9  87.803325 202435.769155 2.137460e+05       0.000000e+00
    10  92.928554 217846.583663 2.317539e+05       2.910383e-11
    11  87.861083 204514.004451 2.173115e+05       2.910383e-11
    12 117.118817 216141.443874 2.453466e+05       8.731149e-11
    13  93.634919 206621.324087 2.247055e+05      -2.910383e-11
    14 127.689875 202484.107982 2.429676e+05       0.000000e+00
    15 165.844461 203719.147456 2.711654e+05       0.000000e+00
    16 168.588508 203214.715404 2.778276e+05      -5.820766e-11
    17 233.843810 209438.433006 3.322003e+05      -5.820766e-11
    18 233.270801 208913.878044 3.382630e+05       5.820766e-11
    19 304.303858 220377.252780 4.225049e+05       0.000000e+00
    20 301.189681 220927.270183 4.297778e+05       0.000000e+00
```

- **How Bias and variance changes**

  From the above graph, it is evident that the bias^2 value has taken a sudden fall when the degree changed from 2 to 3. And so the variance value started increasing when the bias is decreasing. The total error has encountered its lowest value in this 2-3 range. Please note that the y-axis is on the scale of 1e6 as mentioned above. And by the label error, we mean that it denotes the value of corresponding variance, bias2, total error at that degree.

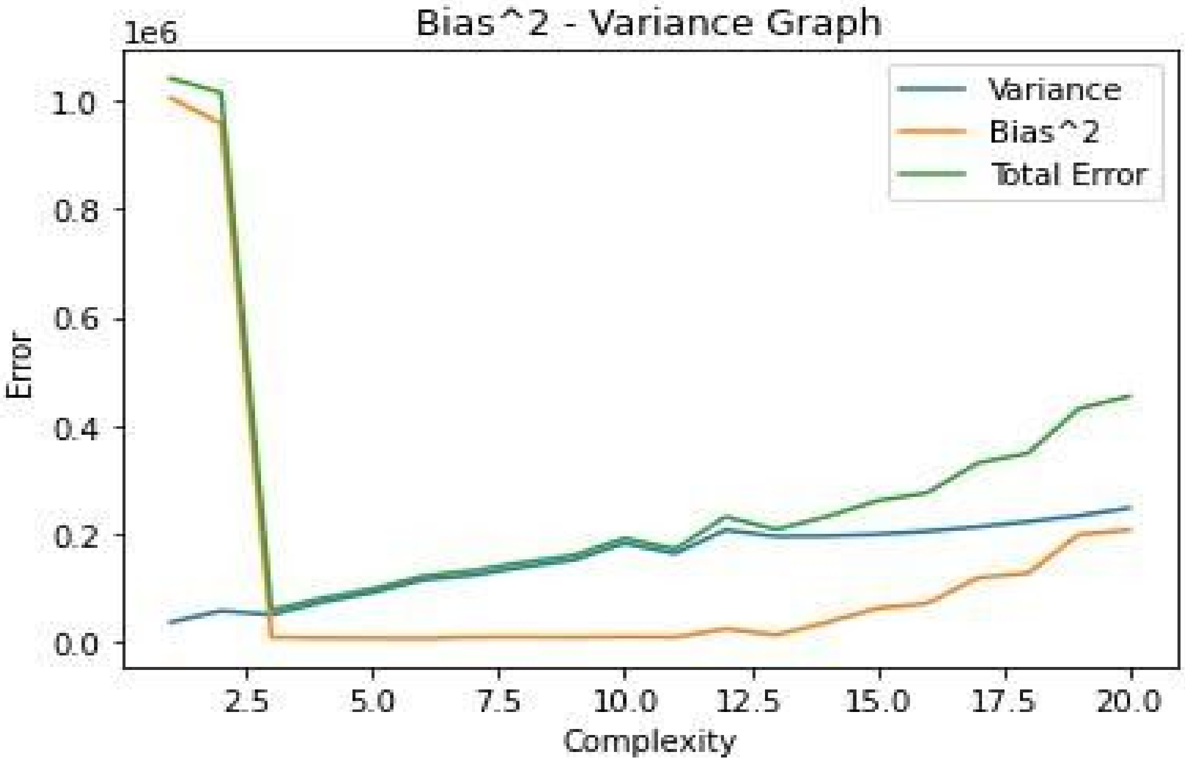
TASK3:



```
Degree           MSE  Irreducible Error
     1 1.034096e+06       0.000000e+00
     2 9.979326e+05       1.164153e-10
     3 3.650776e+04      -7.275958e-12
     4 7.305150e+04       1.455192e-11
     5 7.927119e+04       2.910383e-11
     6 8.718988e+04      -1.455192e-11
     7 1.054323e+05      -2.910383e-11
     8 1.201313e+05      -1.455192e-11
     9 1.378109e+05       0.000000e+00
    10 1.355708e+05       0.000000e+00
    11 1.398416e+05      -5.820766e-11
    12 1.715650e+05       2.910383e-11
    13 1.485088e+05       0.000000e+00
    14 1.598737e+05       2.910383e-11
    15 1.908939e+05       2.910383e-11
    16 2.048232e+05       0.000000e+00
    17 2.598672e+05       0.000000e+00
    18 2.764684e+05      -5.820766e-11
    19 3.572362e+05      -5.820766e-11
    20 3.769174e+05      -5.820766e-11
```

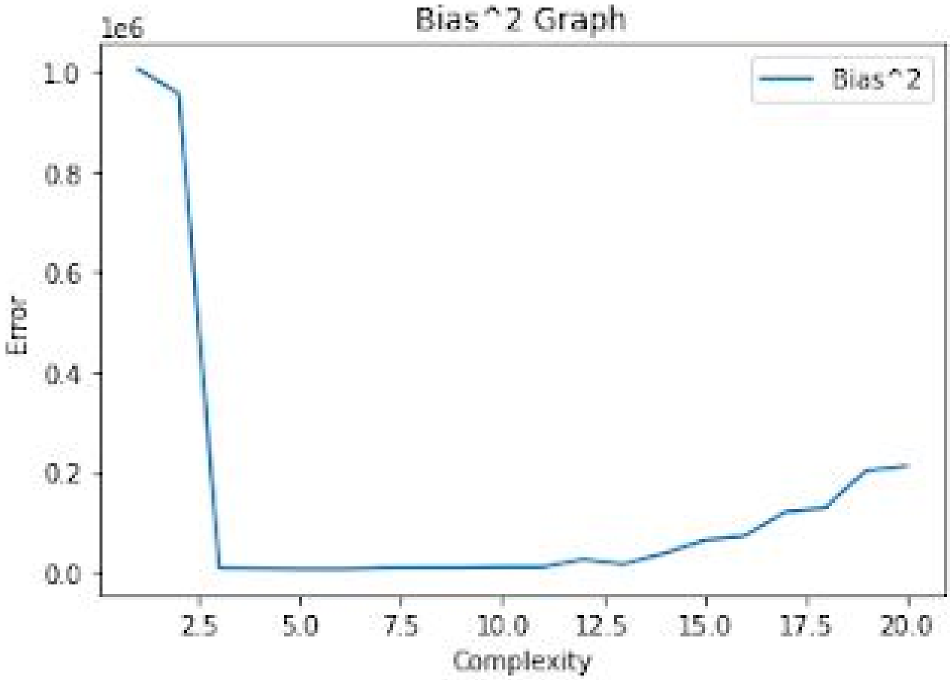
The Irreducible error should not change in theory throughout the degrees as it represents the noise in data but independent of polynomial degree. Here in our above table, the value of IRE is almost in the order of 10^-11 or we can say zero. But there are some instances where the IRE value is zero it may be because the data provided is not real-time, and here the IRE changes maybe because of the method of calculation of MSE, bias^2, variance, and the floating-point errors of python computation.

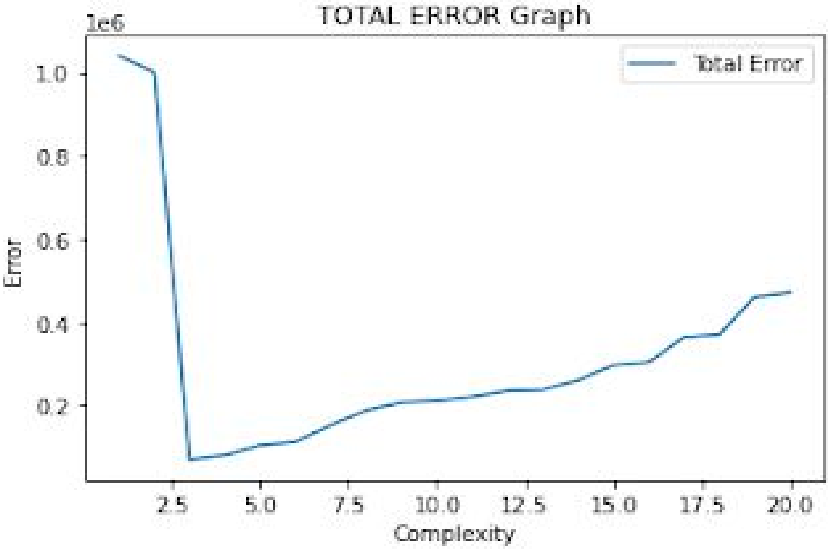
TASK4:

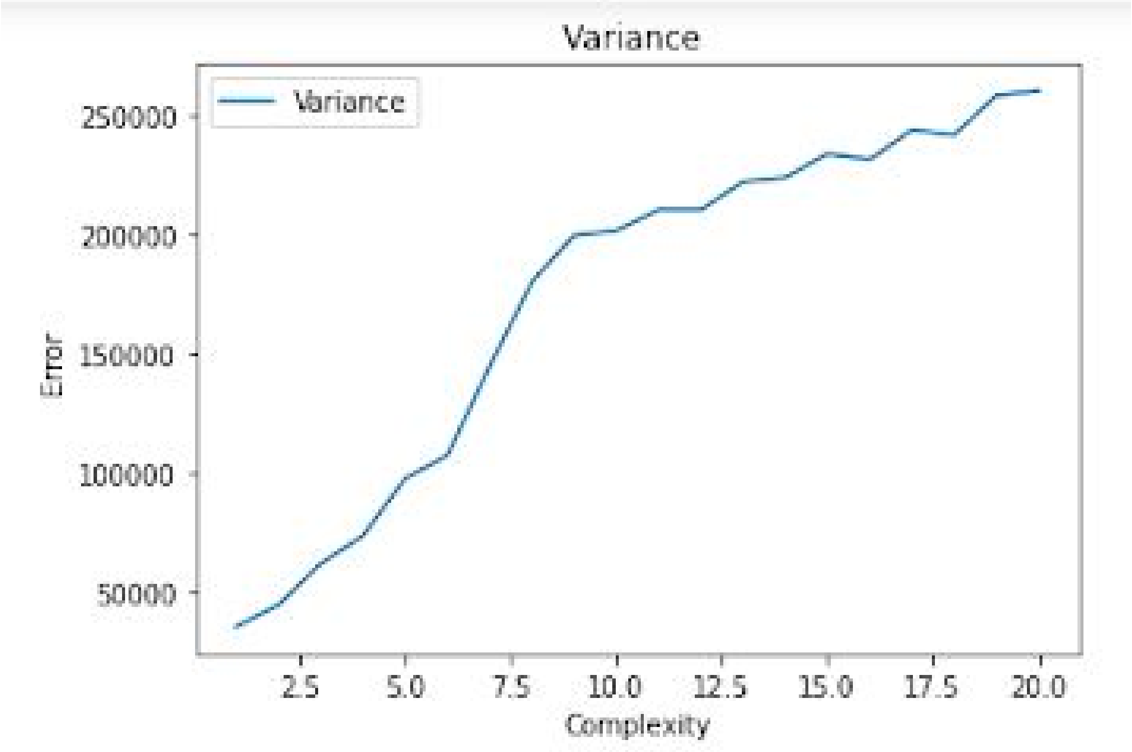


As explained in the bias-variance graph on task 2 the bias when the model complexity is < 2.5 is higher so it is underfitting the data for the polynomials of degree <=2. The bias^2 and variance graphs intersect between 2.5 and 3.5. SO the point between those two has the minimal error so it is the optimal model complexity. When the complexity crosses this optimal point the bias value is low, variance increases and so the model started overfitting the data.

Images :


Bias^2 Graph
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Error
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Complexity
Bias^2


TOTAL ERROR Graph
1e6
1.0
0.8
0.6
0.4
0.2
Error
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Complexity
Total Error



Team Members:

Swamy Naidu CH,	**2019115007**
Narain Sreehith,	**2019101116**