Assignment-1

Team No.: 97
Team Name: Team Caffeine
Team Members: Gautam Ghai (2020101020), Lavisha Bhambri (2020101088)

Task-1 : LinearRegression().fit()

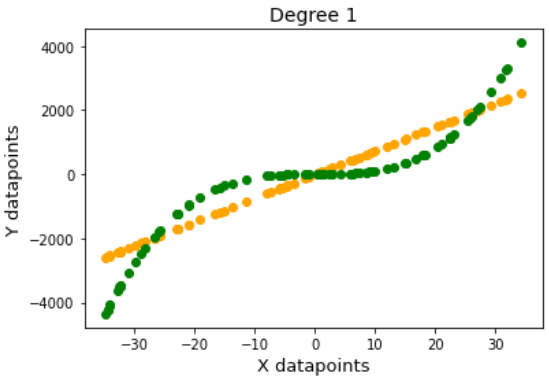
The LinearRegression() in sklearn module creates a linear prediction model by selecting a set of coefficients

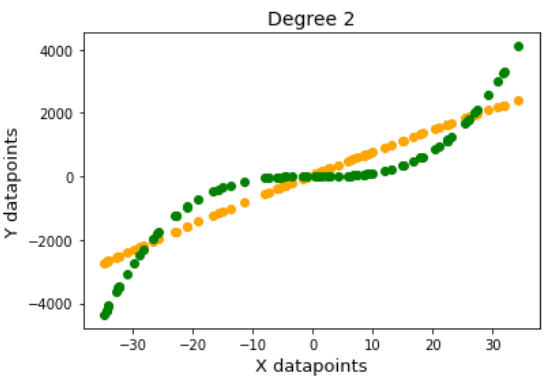
$$w=(w1,...,wn)$$

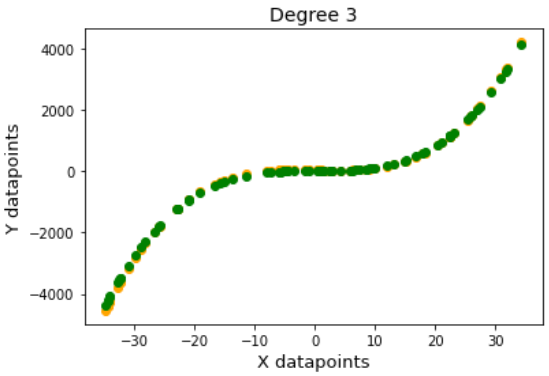
and a bias b to minimise the residual sum of squares between the actual targets in the training dataset, and the targets predicted by linear approximation of the features.
LinearRegression() takes a few parameters which help to decide certain features of the linear model. The LinearRegression().fit() is the function which finds the optimal values of the intercepts where the arguments are the existing input(training dataset). It creates the model to fit the training dataset. The coefficients are initialised by the fit() method to the most optimum values by minimising the Mean Squared Error between the training data and the predicted data. The .fit() function fits any instance of linear regression.

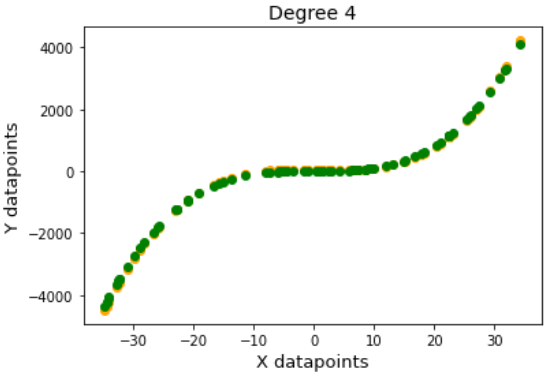
Task-2 : Calculating Bias and Variance

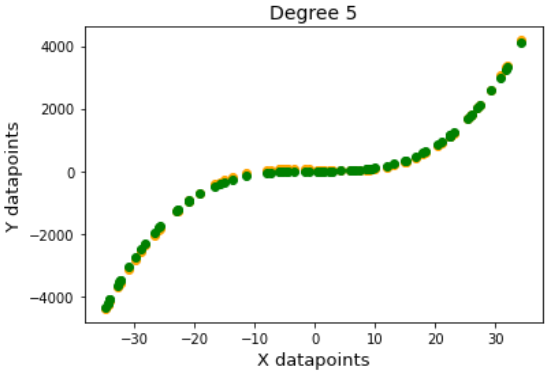
The following graphs depict the predicted values against the original testing values. As can be seen from the following graphs, the worst overlap for predicted and original values is for polynomials of degrees 1 and 2 while it is the best for polynomials with degrees 3-5. So our prediction is that the function most likely represents a polynomial of degree 3 or 4 or 5.

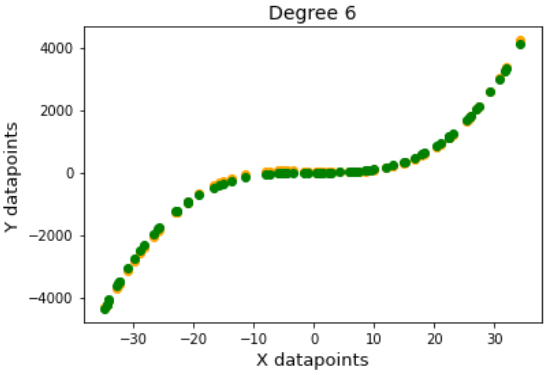

Degree 1
Y datapoints
X datapoints
4000
2000
0
−2000
−4000
−30
−20
−10
0
10
20
30

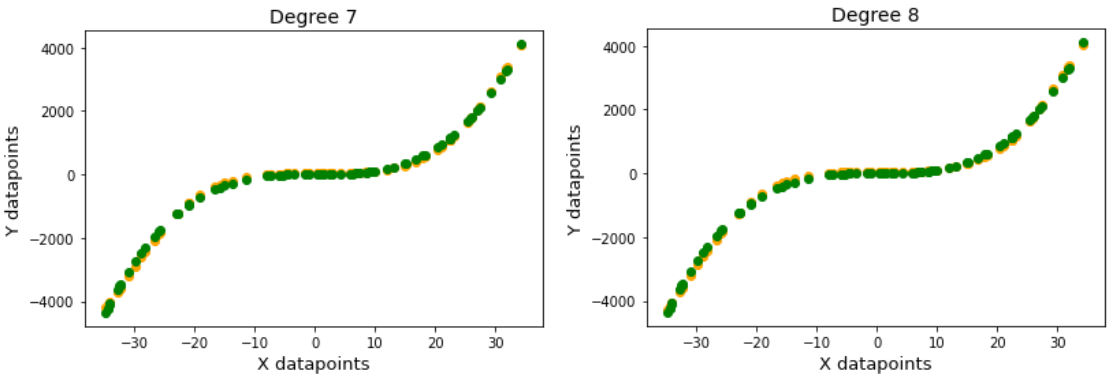

Degree 2
Y datapoints
X datapoints
4000
2000
0
−2000
−4000
−30
−20
−10
0
10
20
30


Degree 3
Y datapoints
X datapoints
4000
2000
0
−2000
−4000
−30
−20
−10
0
10
20
30

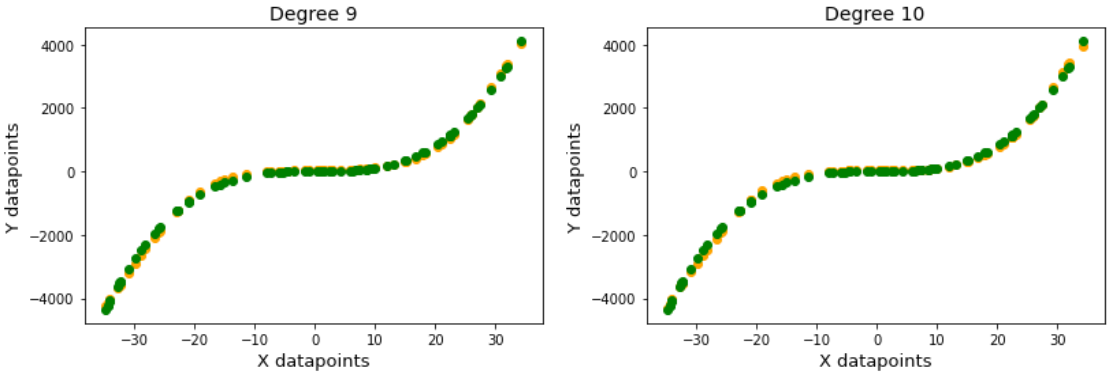

Degree 4
Y datapoints
X datapoints
4000
2000
0
−2000
−4000
−30
−20
−10
0
10
20
30


Degree 5
Y datapoints
X datapoints
4000
2000
0
−2000
−4000
−30
−20
−10
0
10
20
30

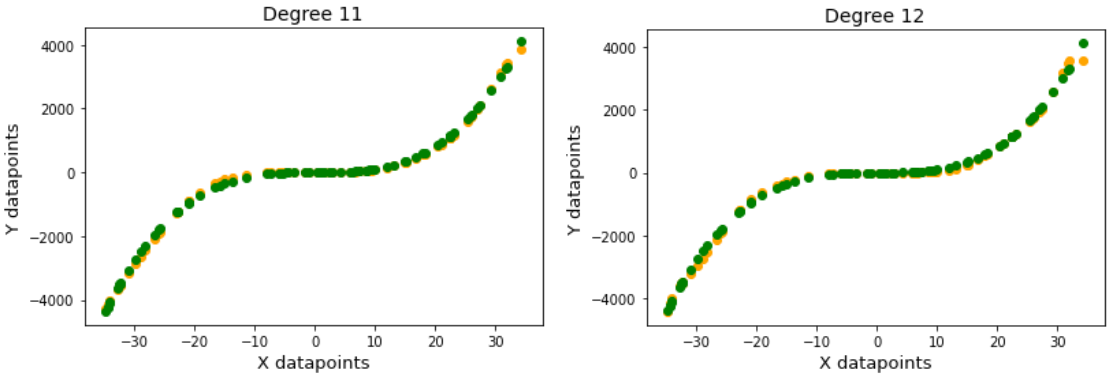

Degree 6
Y datapoints
X datapoints
4000
2000
0
−2000
−4000
−30
−20
−10
0
10
20
30


Degree 7
Y datapoints
X datapoints

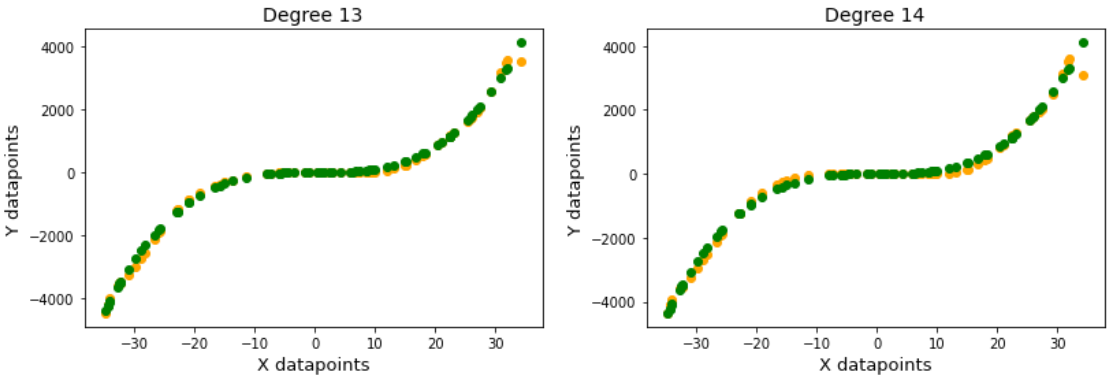
Degree 8
Y datapoints
X datapoints


Degree 9
Y datapoints
X datapoints

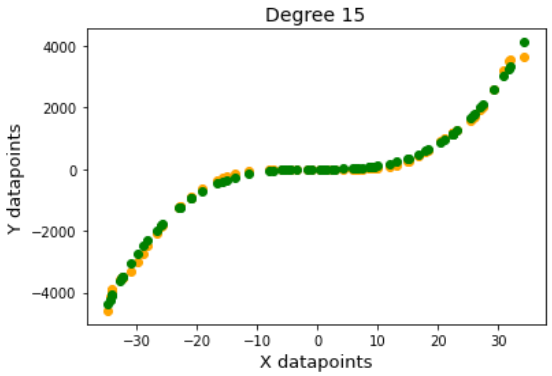
Degree 10
Y datapoints
X datapoints


Degree 11
Y datapoints
X datapoints

Degree 12
Y datapoints
X datapoints


Degree 13
Y datapoints
X datapoints

Degree 14
Y datapoints
X datapoints



If we observe the bias and variance for polynomials from degree 1 to 15 we notice that the bias is high for degrees 1 and 2 (which implies underfitting) and then it abruptly falls for degree=3 and then it remains nearly constant for a while until it starts to slowly increase for higher degrees.

As for variance, it keeps on increasing from degree 1 (where we have an underfitting model) to 15 (where we have an overfitting model). We need to find the right tradeoff between variance and bias to avoid both underfitting and overfitting. From the following table we can infer that the right tradeoff will be for degrees 3 or 4 or 5.

	Bias	Variance
1	573.661	18236.407
2	566.883	33505.379
3	52.295	52186.360
4	58.606	81883.863
5	57.690	89558.200
6	56.866	104817.599
7	56.505	122360.861
8	57.271	140113.264
9	59.857	150808.717
10	62.394	166866.657
11	50.465	146770.817
12	60.184	141791.357
13	59.123	159486.166
14	86.968	130600.389
15	65.405	157613.592

Task-3 : Calculating Irreducible Error

Irreducible error is the error that cannot be reduced by creating good models. It is a measure of the amount of noise in the data.

$$E[(f(x) - \hat{f}(x))^2] = \text{Bias}^2 + \sigma^2 + \text{Variance}$$

$$\sigma^2 = E[(f(x) - \hat{f}(x))^2] - (\text{Bias}^2 + \text{Variance})$$

where $f(x)$ represents the true value,
$\hat{f}(x)$ represents the predicted value,
$[E[(f(x) - \hat{f}(x))^2]$ is the mean squared error (MSE) and
σ^2 represents the irreducible error.

For almost all cases from degree 1 to 15, the irreducible error remains close to 0. Ideally irreducible error should be 0 but some small amount of error arises due to the presence of noise in virtually every data.

	Degree	Irreducible Error
1	1.0	2.328306e-10
2	2.0	-3.492460e-10
3	3.0	-1.455192e-10
4	4.0	2.037268e-10
5	5.0	-8.731149e-11
6	6.0	-1.891749e-10
7	7.0	8.731149e-11
8	8.0	2.328306e-10
9	9.0	5.820766e-11
10	10.0	-2.910383e-11
11	11.0	2.910383e-10
12	12.0	1.164153e-10
13	13.0	8.731149e-11
14	14.0	2.910383e-11
15	15.0	0.000000e+00

Task-4 : Plotting Bias^2 – Variance graph

We observe that the bias drops suddenly from degree=2 to degree=3 then remains about the same and then for higher degrees it starts increasing gradually. As for the variance it gradually increases from degree 1 to 15 and the error drops steeply from degree 2 to 3 and then increases. In short the model moves from underfitting to overfitting with the minimum error at degree=3. Thus we conclude that the data is best fitting for a model with degree=3 as the error is the least and it shows that a good tradeoff exists between bias and variance.