Assignment 1

Team Number: 64
[Dhruvee Birla: 2019115008]
[Pahulpreet Singh: 2019113013]

Task 1

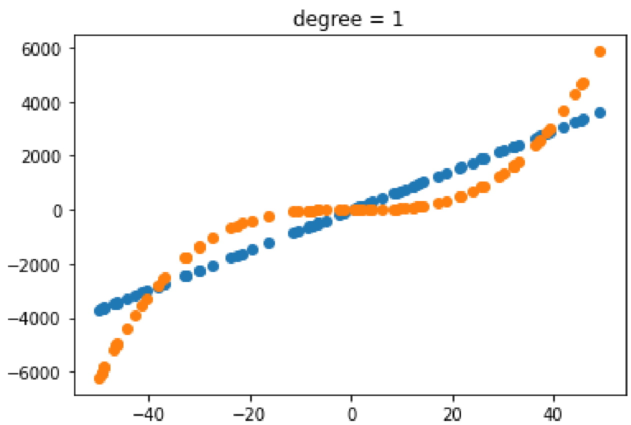
A linear regression model represents a linear equation which has a certain set of inputs (x) and predicted outputs (y) for those inputs.

The equation gives each input value a one unit factor which is called coefficients. The intercept or the bias coefficient is another coefficient which gives the line freedom of movement.

$$\mathbf{y = B_0 + B_1(x)}$$

Here, B_0 is the bias coefficient/intercept

The intercept is the expected mean value of Y when all X (predictor of the regression equation) equals zero.

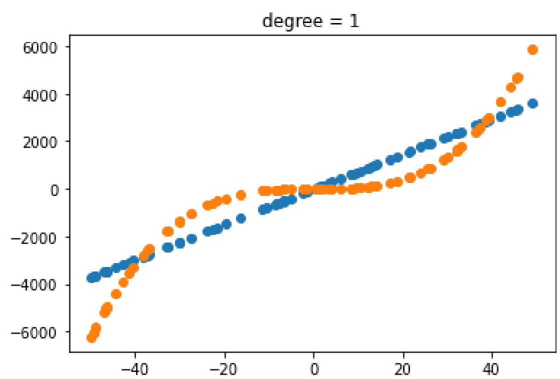


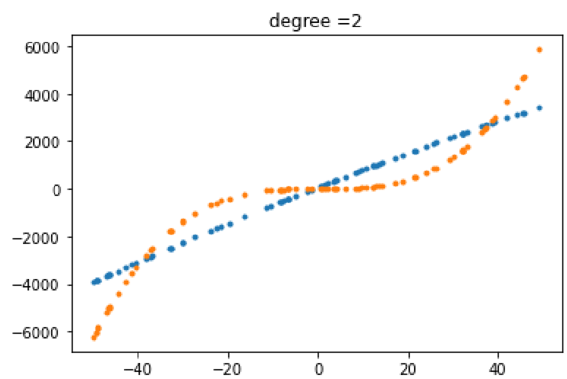
.fit() function finds the optimal values of the intercepts where the arguments are the existing inputs and outputs. An estimator is fit onto the model to predict the outputs of the unseen inputs. A classifier is an estimator instance which is fitted to the model and its purpose is to learn from the model. The .fit() function fits any instance of the Linear Regression.

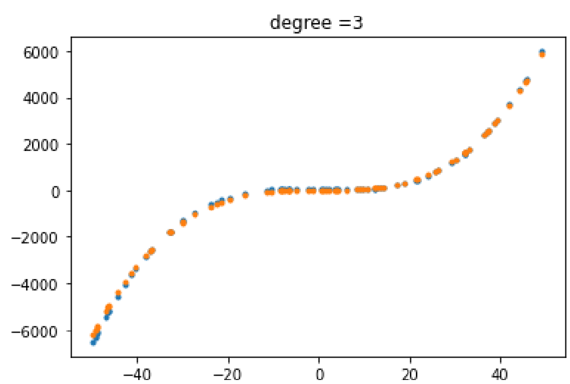
Task 2

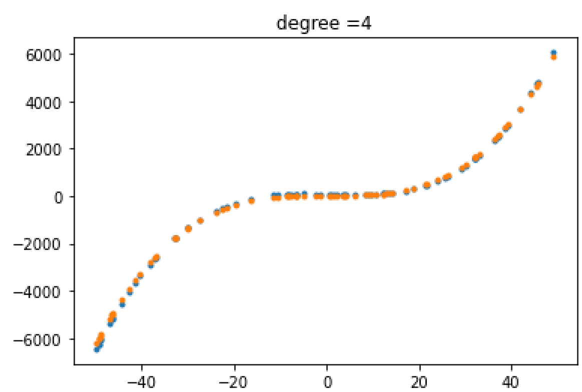
Below are the graphs depicting the predicted values against the original testing values, that is, the relationship between predictor and response. Therefore, a complete overlap would mean similar values of predictor and response which would be the ideal case.

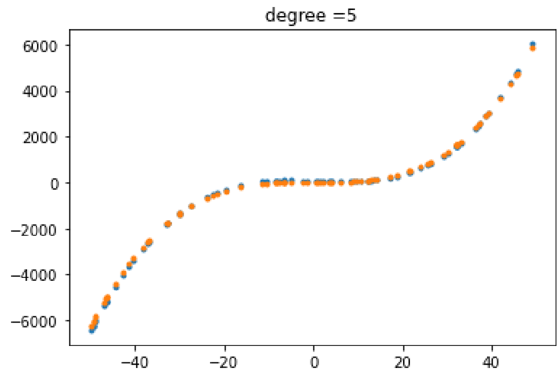
As can be interpreted from the graphs below, the functions with degree 1 and degree 2 do not have overlapping values for predicted and original values whereas functions with degree 3 to degree 5 have the best overlap. Hence, our assumption is that our data represents a function of either degree 3,4 or 5.

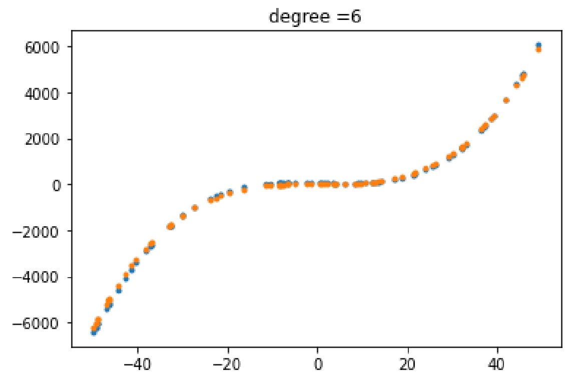


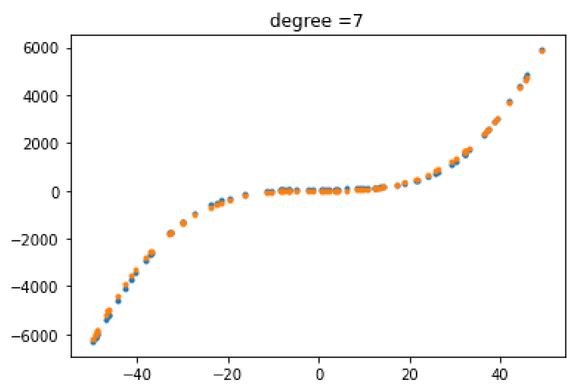


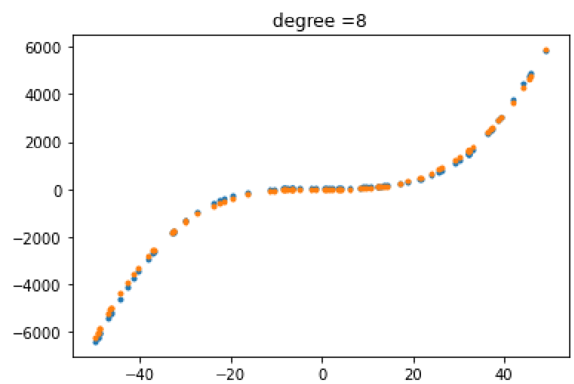


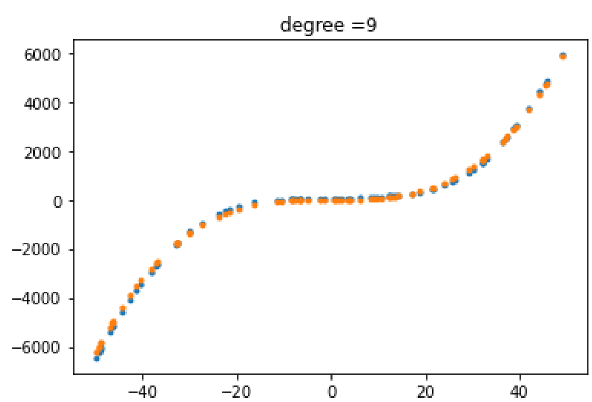


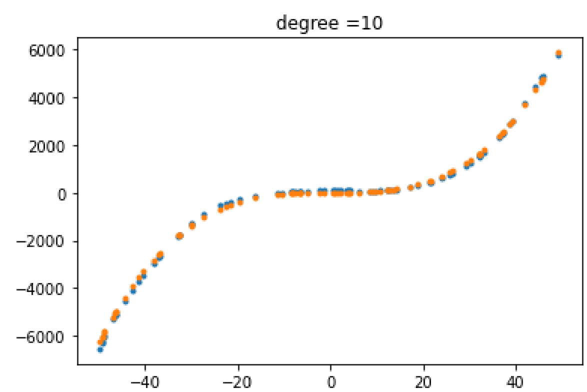

degree =5
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

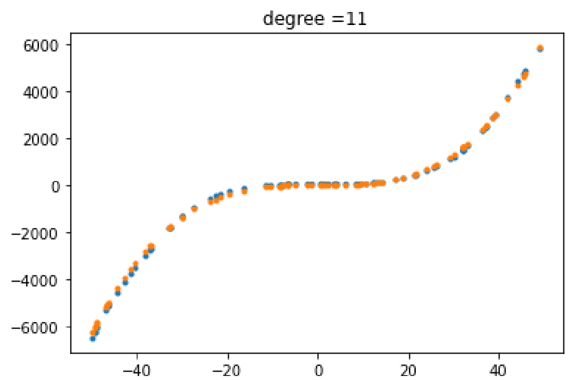

degree =6
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

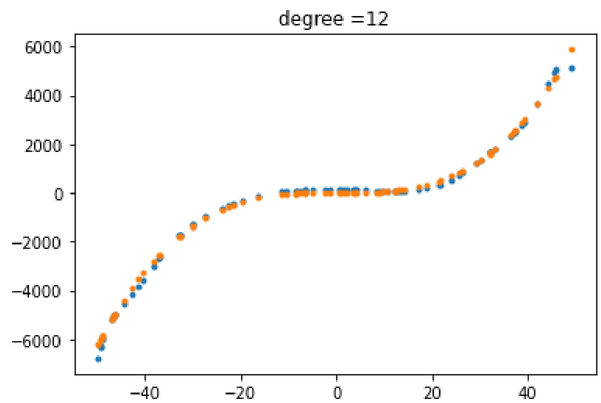

degree =7
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

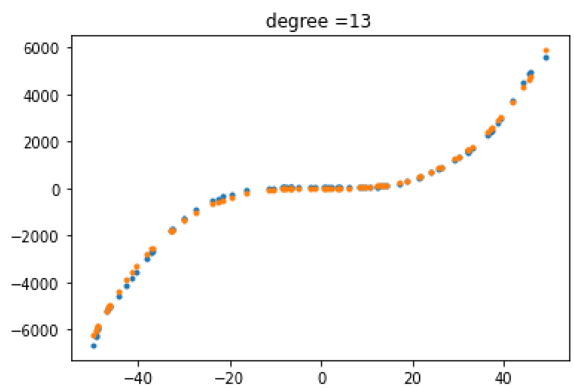

degree =8
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

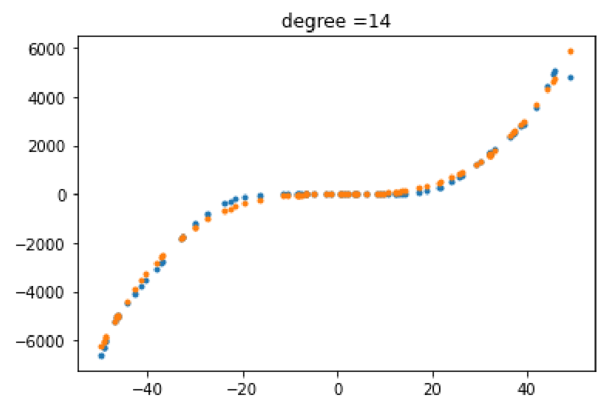

degree =9
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

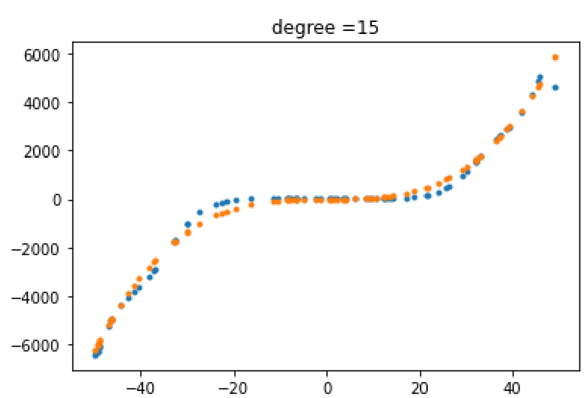

degree =10
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

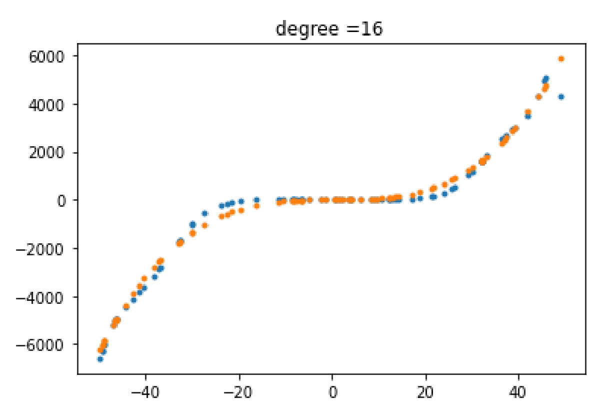

degree =11
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

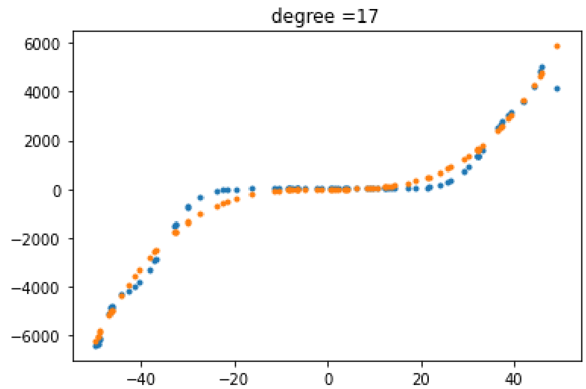

degree =12
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

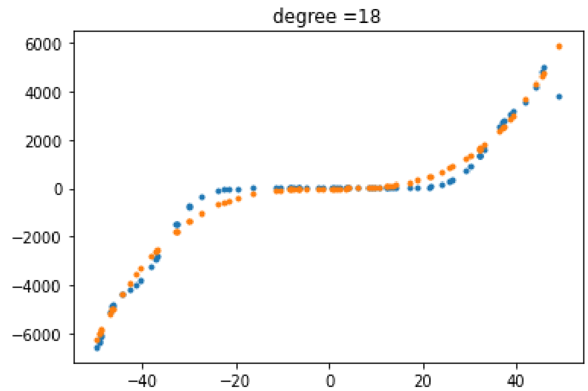

degree =13
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

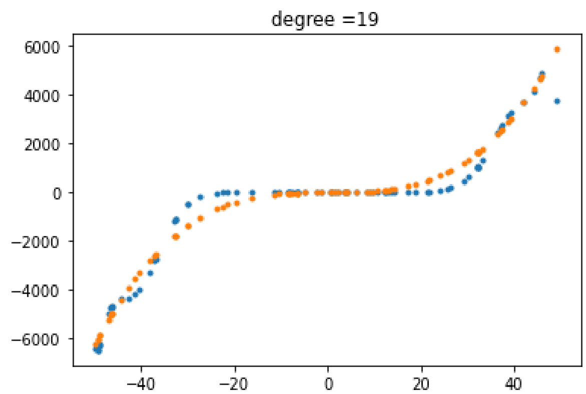

degree =14
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40

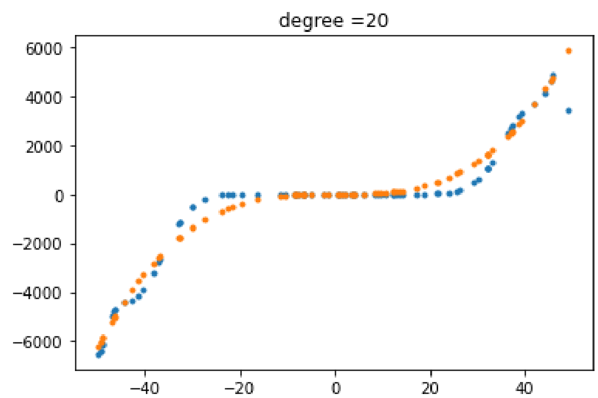

degree =15
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40


degree =16
6000
4000
2000
0
−2000
−4000
−6000
−40
−20
0
20
40









In contrast to variance, bias does not change a lot across the function classes. The value of bias initially decreases, the more or less remains constant and starts increasing again. It starts decreasing from degree 2 and starts increasing from degree 8. Variance, on the other hand, keeps increasing from degree 1 to degree 20.
An increase in variance indicates a better fit of the model with every change in function classes. It indicates a move from an underfitting model to an overfitting model. A good fit of the model would be when there is balance between bias and variance.

We should prefer a tradeoff between bias and variance rather than an underfitting or an overfitting graph. Therefore, from the tabulated values of bias and variance, we infer that an almost best fit would be for a function with degree 3 or degree 4. In the previous subsection, we also concluded that a good fit seems to be taking place for functions with degree 3, 4 and degree 5.

In conclusion, the best fit model is either with function of degree 3 or degree 4.

	Bias	Variance
1	1002.367	38342.711
2	976.842	49402.325
3	95.294	57686.915
4	93.400	69606.637
5	87.749	84091.408
6	87.956	112100.679
7	92.049	118728.980
8	92.810	137807.376
9	92.967	156679.727
10	100.504	146150.711
11	99.346	161003.971
12	149.505	141256.453
13	126.207	172510.959
14	189.188	150702.772
15	252.217	172432.757
16	271.807	175326.617
17	350.885	193007.226
18	368.345	202425.499
19	457.996	214840.258
20	475.176	231621.403

Task 3

Irreducible error is the error that can't be reduced however good the model is implying that irreducible error is independent of the model. It is the amount of noise in the model, that is, it is the error that arises when X cannot provide information for Y, it could be because of unmeasured variables which were useful in predicting Y, unmeasurable variation, etc.

$$Y = f(X) + e$$

Where X is the input variable,Y is the response variable and e is error

For a good model, there should be a good balance between bias and variance for a minimal total error.

$$\textbf{Total error} = \textbf{Bias}^2 + \textbf{Variance} + \textbf{Irreducible error}$$

The irreducible error remains more or less the same, that is, close to zero across all the function classes.

	Irreducible error
1	0.000000e+00
2	0.000000e+00
3	0.000000e+00
4	0.000000e+00
5	-1.455192e-11
6	-1.455192e-11
7	0.000000e+00
8	-2.910383e-11
9	0.000000e+00
10	-2.910383e-11
11	0.000000e+00
12	0.000000e+00
13	2.910383e-11
14	5.820766e-11
15	-2.910383e-11
16	5.820766e-11
17	1.164153e-10
18	5.820766e-11
19	-5.820766e-11
20	-5.820766e-11

Task 4

The value of bias decreases and that of variance increases between the values 1 and 3. We can also observe from the graph below that the error is the lowest at X = 3.

The goal of any machine learning algorithm is to have low bias and variance. Therefore, a good model is one with a good balance between bias and variance or with a good tradeoff between bias and variance that it minimizes the total error.

Before X = 3, bias is significantly high and drops suddenly from around X = 2 and hits its lowest at X = 3. Since, before X = 3, we have a high bias and low variance (variance is lowest at X = 1), this points towards an underfitting model as hypothesized in Task 2.

After X = 3, variance increases slowly and bias remains constant till X = 11 and then increases. Therefore, after X = 3, the bias is low and the variance is comparatively high till before X = 19, which points towards an overfitting graph.

The graph below is in sync with our observations in Task 3, that is, the model moves from an underfitting to an overfitting model. Since, the error is the lowest at X = 3 which implies a good balance exists between bias and variance at X = 3.

In conclusion, the data is best fitting for a model of degree 3.
This suggests that *Y = f(X)* is cubic.