MDL Assignment 1 Report

Tanvi Narsapur(2019111005)
Ahana Datta (2019111007)

Task 1:
Write a brief about what function does the method, LinearRegression().fit() performs.
Ans. Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories.

The LinearRegression.fit() method trains or fits the data.
LinearRegression().fit(X,Y)-> puts the x values and y values in the given function respectively. Where the function is a polynomial.
Thus for the function of the form

$$y = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

the LinearRegression.fit() method predicts the most optimal values $(a_n, a_{n-1}, \dots, a_1, a_0)$ given the values of x and y. The method trains the model over the training dataset so that it can be used to predict the values for the test dataset.

Task 2:

Degree	Variance	Bias
1	26724.17908	820.228845
2	60428.39375	810.53006
3	76775.15439	67.92933
4	92460.55232	74.659798
5	131176.5842	76.991013
6	146850.0695	72.789141
7	164084.2248	79.788973
8	184954.2453	82.090933
9	196923.7579	78.29139
10	187474.9855	84.400848
11	215046.2794	76.021089
12	211462.4483	104.648186
13	219104.6389	80.922508
14	221004.4144	111.150452
15	219878.2786	153.469792
16	231507.1764	157.671967
17	224297.4135	230.29227
18	238078.756	231.938917

19	232756.3627	300.926429
20	248928.4888	301.011102

The minimum value of Variance is observed for degree = 1.
The minimum value of Variance is observed for degree = 3.

Observing the values of Bias and Variance obtained for each function class, a particular trend is observed in both these values.

In general, the value of Variance is observed to have an increasing pattern with an increase in model complexity.

For the Bias value initially, a sharp fall is observed from degree 2 to 3 and the value remains almost constant As we go on increasing the value of the degree of the function. For the function classes having a value of degree closer to 20 the bias value is observed to have an increasing trend.

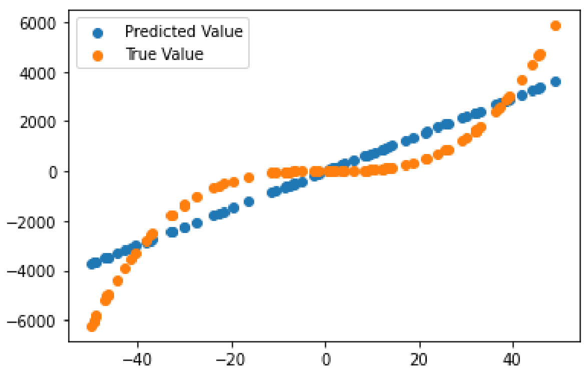
For the models with function class as lower degree polynomial, we have not extracted enough information from the training dataset and so the model is underfitting. Because of this, the predicted values do not match with the test dataset (this can be observed from graphs for degree 1 and 2 below). The models being underfitted, have high bias and low variance values(as seen in the above table).

As the degree of the polynomial function increases, the complexity of the model goes on increasing. Thus the model is better trained and fits the test data better. As the model gets trained better, the bias value decreases. Variance increases since the complexity of the model goes on increasing and the model is overfitting. The same can be seen from graphs for degree 3 to 13.

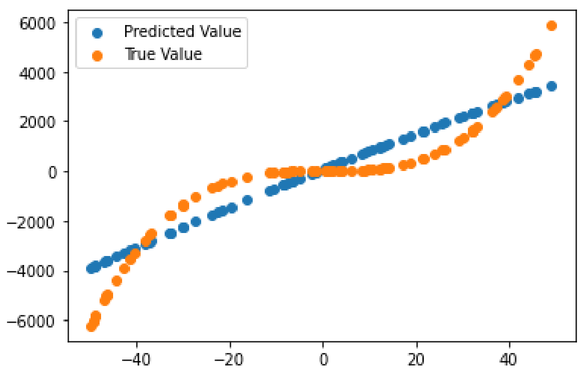
When the degree of the polynomial is higher than 13, the test data does not fit accurately with the predicted values (the graphs for degrees higher than 13 do not have a good overlap for true values and predicted values). Thus the value of bias increases.

Below are the graphs showing predicted value and true value for each point in the test data set for every degree of the class function (from 1 to 20)

Degree 1:

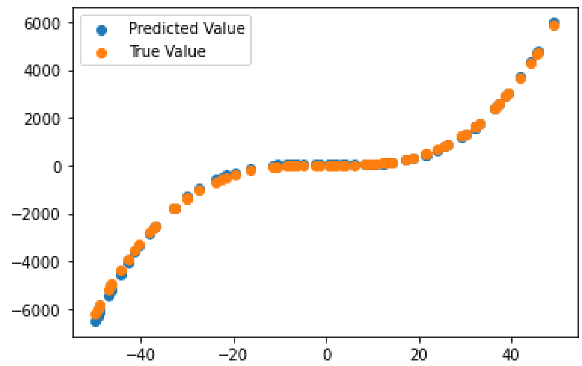


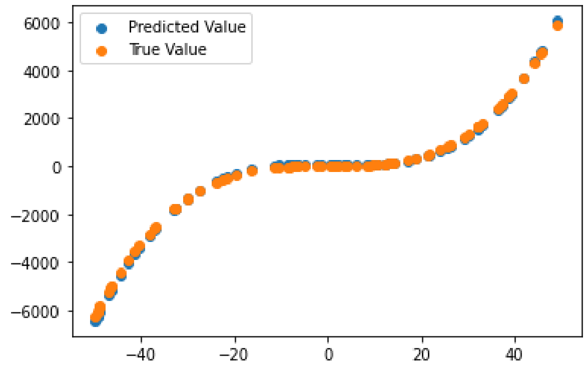
Degree 2:



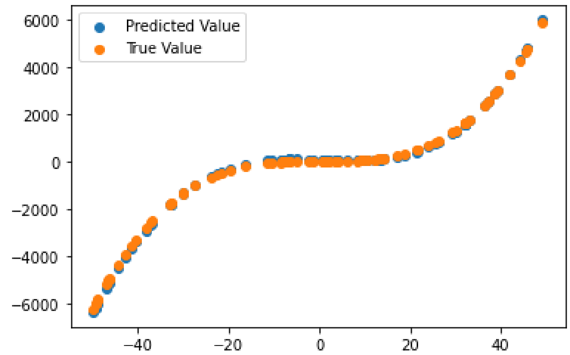
Degree 3:

Degree 4:

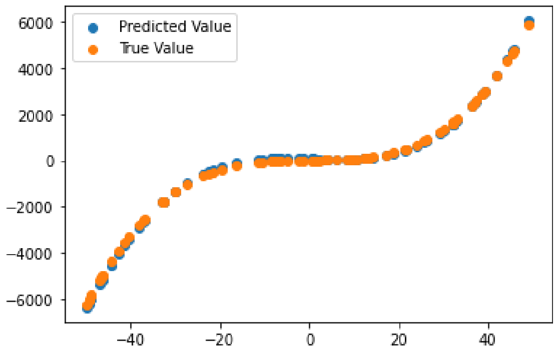




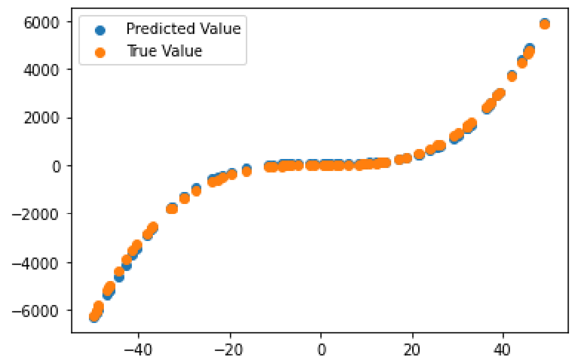
Degree 5:



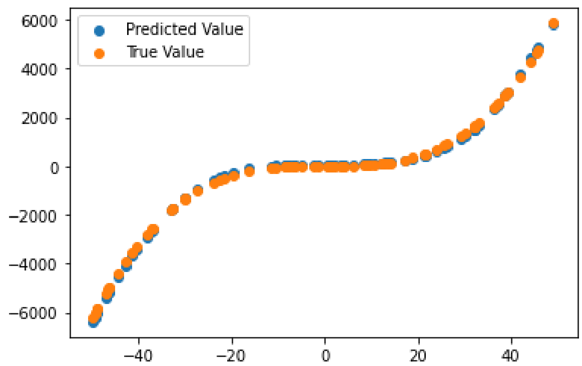
Degree 6:



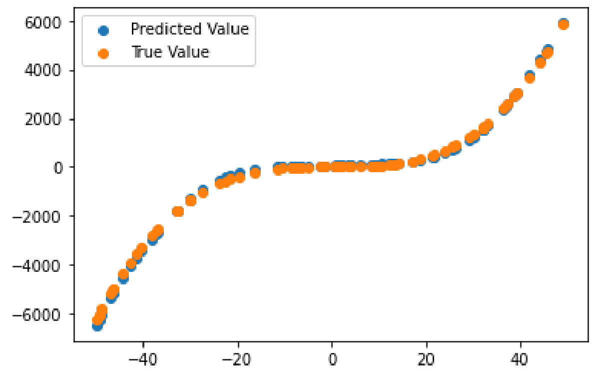
Degree 7:



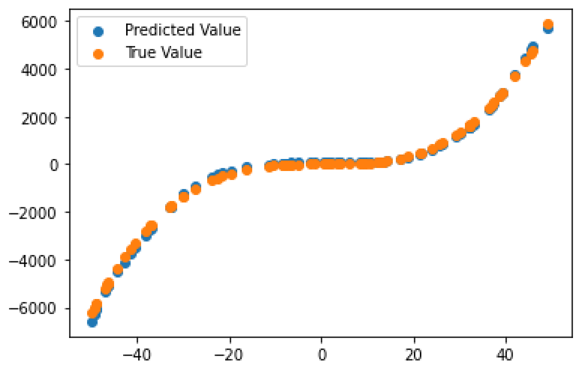
Degree 8:



Degree 9:

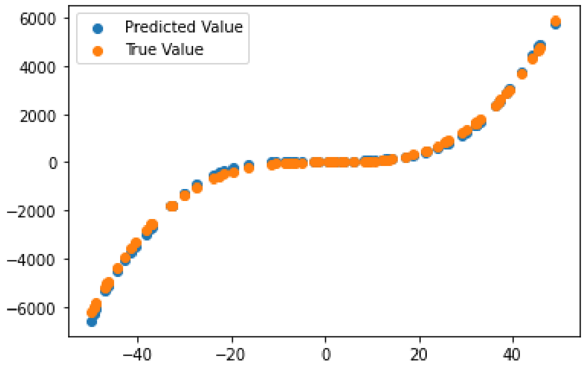


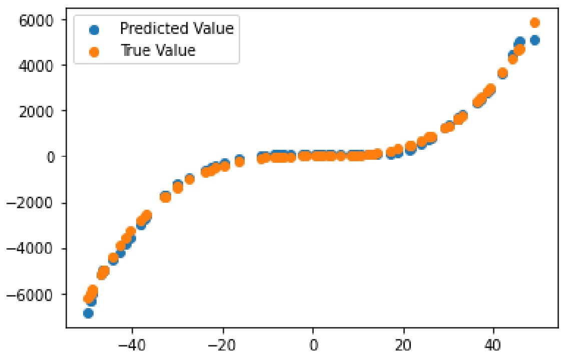
Degree 10:



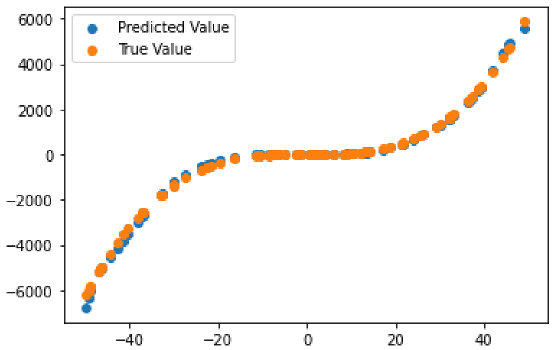
Degree 11:

Degree 12:

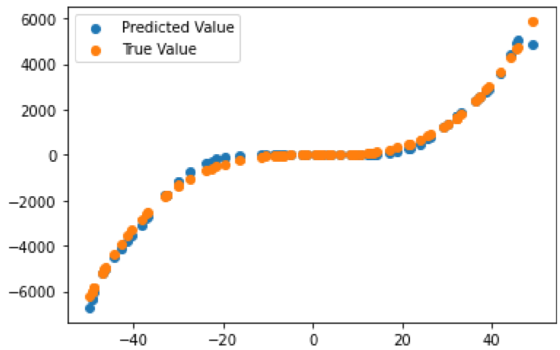




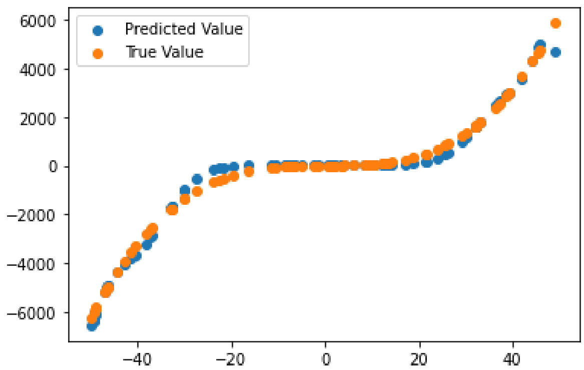
Degree 13:



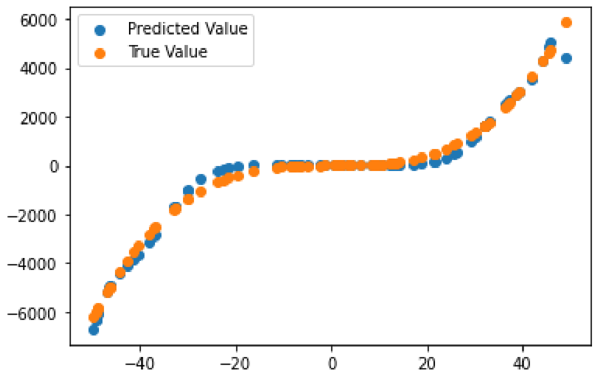
Degree 14:



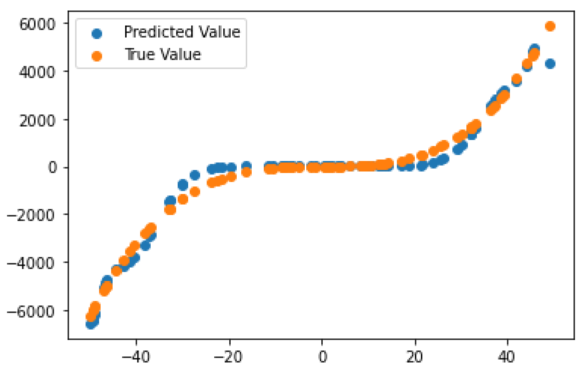
Degree 15:



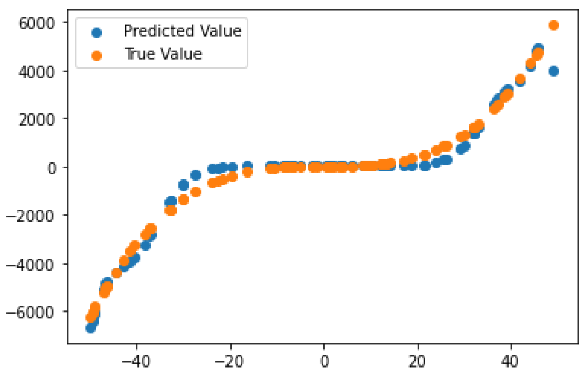
Degree 16:



Degree 17:

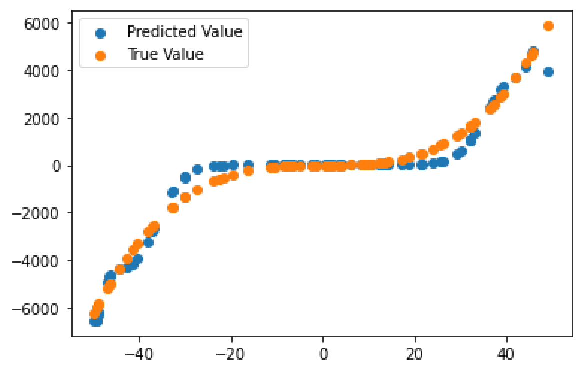


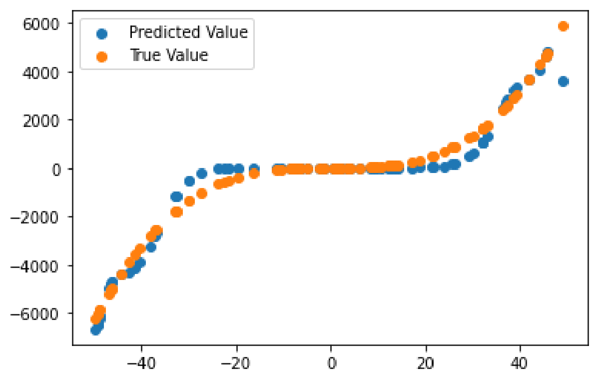
Degree 18:



Degree 19:

Degree 20:





Task 3:

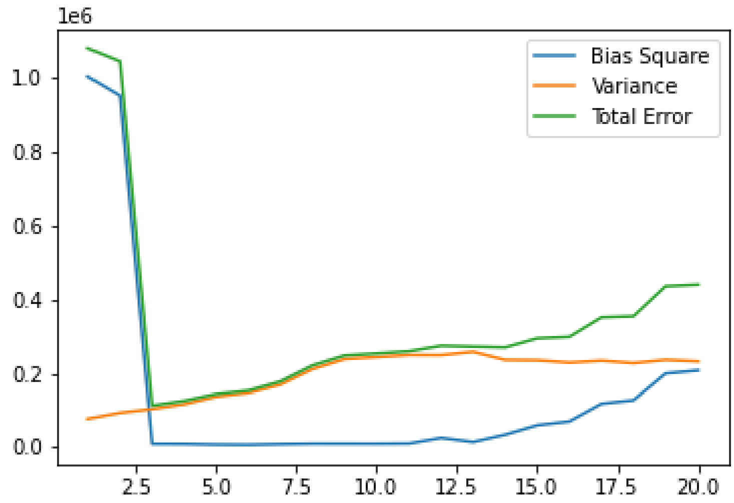
Degree	Variance	Bias	Irreducible error	Total error
1	26724.17908	820.228845	-1.05e-10	1.03e+06
2	60428.39375	810.53006	4.64e-11	1.02e+06
3	76775.15439	67.92933	-5.37e-12	8.56+04
4	92460.55232	74.659798	7.64e-12	1.00e+05
5	131176.5842	76.991013	-3.27e-12	1.40e+05
6	146850.0695	72.789141	-1.22e-11	1.54e+05
7	164084.2248	79.788973	2.16e-11	1.73e+05
8	184954.2453	82.090933	-1.58e-11	1.94e+05
9	196923.7579	78.29139	-1.11e-11	2.05e+05
10	187474.9855	84.400848	5.18e-12	1.97e+05
11	215046.2794	76.021089	9.55e-12	2.23e+05
12	211462.4483	104.648186	4.00e-12	2.32e+05
13	219104.6389	80.922508	-1.36e-11	2.31e+05
14	221004.4144	111.150452	-1.60e-11	2.50e+05
15	219878.2786	153.469792	3.46e-11	2.75e+05
16	231507.1764	157.671967	5.71e-11	2.95e+05
17	224297.4135	230.29227	2.32e-11	3.37e+05
18	238078.756	231.938917	-4.00e-12	3.58e+05
19	232756.3627	300.926429	2.76e-11	4.27e+05
20	248928.4888	301.011102	4.19e-11	4.51e+05

Irreducible error is the error that can’t be reduced by creating good models. It is a measure of the amount of noise in our data. Here it is important to understand that no matter how good we make our model, our data will have a certain amount of noise or irreducible error that can not be removed.

Irreducible error, as the name suggests, is irrelevant to the underlying model and has to do with the inherent noise in the problem. This noise can represent noise coming from data quality (e.g., inaccuracies in collection or reporting of data), from non-deterministic behavior of the underlying phenomenon, and in general, any type of noise that cannot be easily defined.
From the tabulated values of irreducible error, we can see that the values of irreducible error are extremely small across all models.

Since the irreducible error is inherent to the problem itself, it doesn't vary significantly over different class functions. This is also observed in the values tabulated in the table above - the change in the value of irreducible error is not very significant.

Task 4:
The plot for the values of Bias square and Variance plotted for each degree of the function class polynomial can be given by



From the above graph and tabulated values, we observe that with an increase in the complexity of the model the bias decreases and variance increases.

A machine learning model is said to be underfitting when the model is unable to perform well on the training dataset as well as the test dataset. Underfitting occurs when we are not using enough data to train the model. That is it fails to capture the underlying trend of the data. For an underfitting model, the bias value is high and the variance value is low. This can be seen from the above graph.

When we are extracting too much information from the training dataset, the model becomes overfitted. As too much information is being extracted, the fluctuations in the dataset become prominent. This model will not help us generalize data and derive patterns from them. Thus the model may perform poorly on data sets that have not been seen before. Thus, an Overfitting model has a high variance value, also such a model has a low bias value.

For the given dataset, the model is Underfitting when the functional class is lower degree polynomial. Whereas the model is Overfitting when the functional class is higher degree polynomial.

To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.

Total Error = $Bias^2$ + Variance + Irreducible Error

An optimal balance of bias and variance would never overfit or underfit the model. From our bias-variance tradeoff graph, we can see that model with degree=3 would be a good model. This is because the total error is minimum for degree=3. Therefore the optimal model complexity is when degree=3.