MDL Assignment-1 Report

Task 1 -
The `LinearRegression().fit()` function returns a line object that best models the set of data passed as parameters to it through linear regression. This best fit linear model is created by minimizing the total square error of the training data by the **gradient descent** method.

fit(X, y, sample_weight=None)

Parameters

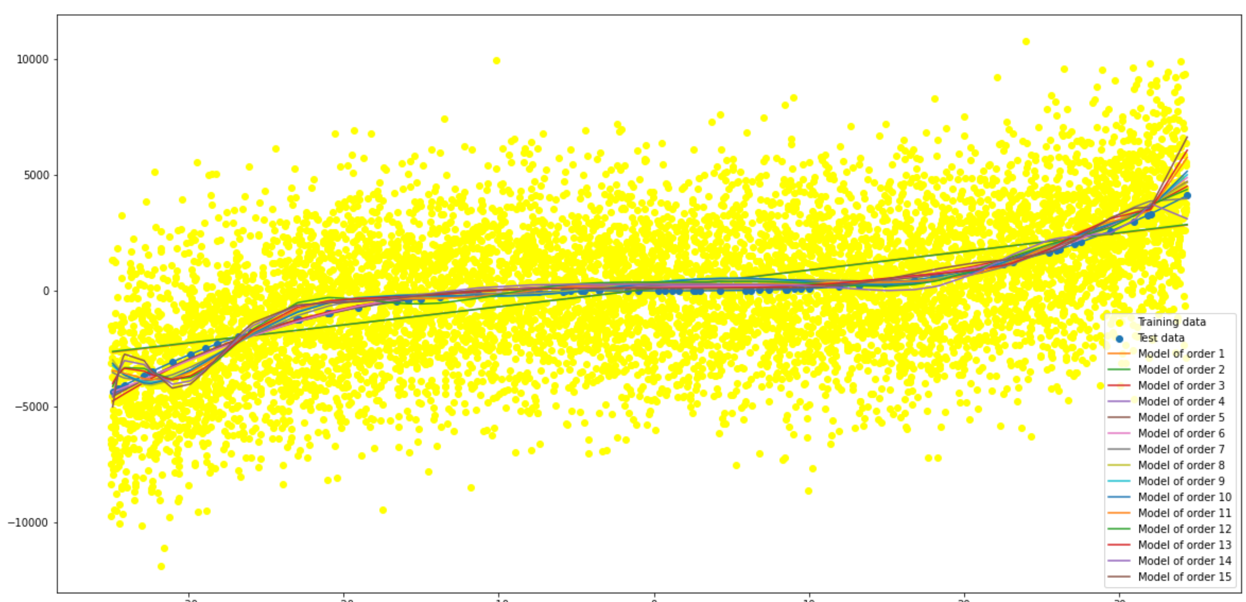
-*X* {array-like, sparse matrix} of shape (n_samples, n_features) Training data.

-*y* {array-like} of shape (n_samples,) or (n_samples, n_targets) Target values. Will be cast to X's dtype if necessary.

-*sample_weight* {array-like} of shape (n_samples,), default=None, Individual weights for each sample.

Returns
selfobject Fitted Estimator.

Task 2 -



Squared-Bias vs Variance:
For calculating bias, variance, and mean squared error, we have to take mean over all estimators of a single polynomial class f^(x) and calculate the values of bias squared, variance and the mean squared error for each point on each estimator with respect to f^(x). Then we take the mean over all these points to obtain single values. As the degree changes, squared bias decreases, and their variance increases.

Bias - It is a measure of the simplicity of our model. As it measures how far our average estimator value is from the actual function output (which is test data y) it can tell us if our model is being 'biased' toward being simple and it does not fit the data well. If we increase the degree of the polynomial which we use as an estimator, bias will eventually decrease because the model would fit data better (and become more complex).

Variance- It is a measure on how far a model which is based on a particular realization of the training data is different from the average estimator of that polynomial degree. In case of lower degree polynomials, all the estimators would be centered (i.e not much different from each other) and thus would have low variance. As the model become more and more complex, the realization of the training data has a bigger impact, and thus the variance increases.

Degree	Bias	Squared bias	Variance
1	163.561	492332	35117.6
2	159.438	468788	53836.6
3	-9.90197	4200.23	67030.1
4	-7.16853	4156.43	88454.5
5	-5.42154	3831.31	98446.5
6	-5.10024	3941.34	129710
7	-2.5518	4344.73	133694
8	-4.2702	4177.67	148721
9	-2.75256	3987.72	165415
10	-3.56527	4305.08	181676
11	-5.08154	4151.92	183554
12	-3.10474	9559.01	208314
13	-11.4694	6131.11	177497
14	-12.1248	23191.9	226636
15	-11.5964	11170.3	197344

The strange rise in bias square as the degree rises is because of heavy overfitting of the higher degree polynomials, and thus their average model could vary even more than the test data (the model overfits the training data and thus it misses the mark on the testing data).

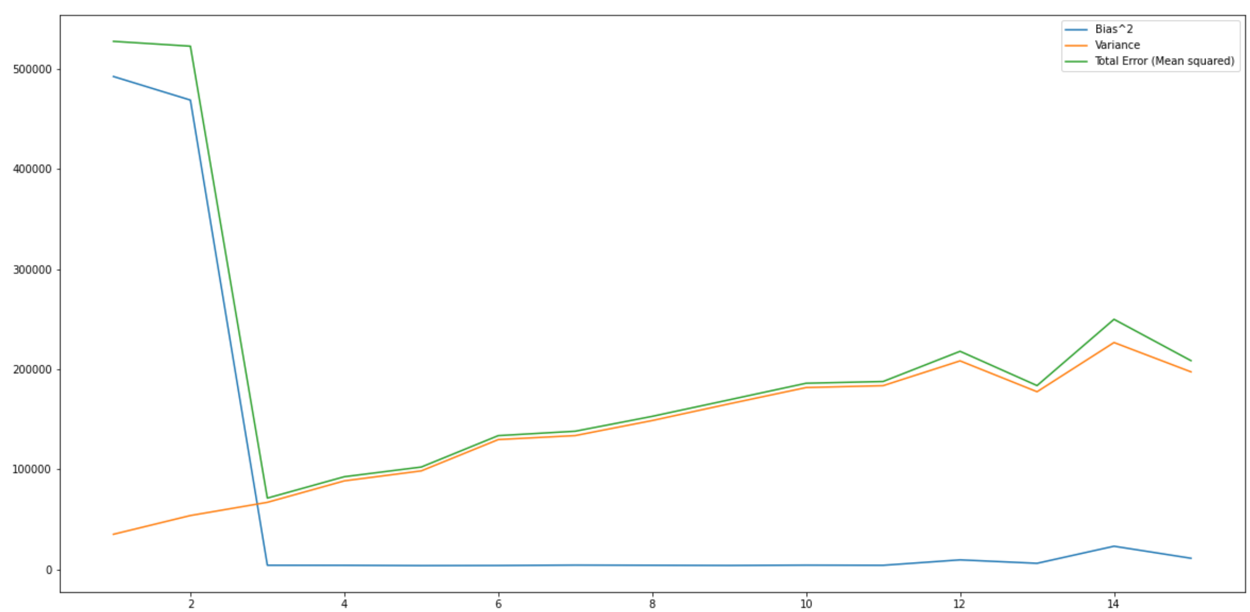
Task 3-

Irreducible Error

From the image below, the irreducible error values change hardly. Also, they are very near to 0 (some are negative due to python precision errors). For realistic purposes, this could be assumed to be 0. The irreducible error does not change,because it is a property of the test data. We are assuming that noise is normally distributed, and our irreducible error is sigma squared. This value would be the same among realizations of the testing data, and across polynomial degrees. Looking at the scatterplot, we can see that the test data (in blue) is exactly fitting a polynomial. This leads us to believe that there was a minimal noise which was introduced in the test data. This also explains us why the observed irreducible error is constant and is very close to 0.

Degree	Irreducible error
1	-7.27596e-11
2	8.73115e-11
3	-1.45519e-11
4	1.45519e-11
5	-1.45519e-11
6	-1.45519e-11
7	0
8	0
9	0
10	-2.91038e-11
11	2.91038e-11
12	0
13	-2.91038e-11
14	0
15	2.91038e-11

Task 4-



From the above Graph, we can say the following:

High Bias, Low Variance -
Whenever a model (polynomial degree) has high bias and low variance, it is 'underfitting'. This can be explained by models with low degrees having high biases. This is as they tend to simplify the estimators which are very similar to each other and thus they result in underfitting.

High Variance, Low Bias -
Whenever a model (polynomial degree) has high variance and low bias, it is 'overfitting'. This can be explained by models having high degrees and high variances. This is because they tend to create complex estimators which vary significantly from each other and result in overfitting.

Ideal model -
The lowest error (mean squared) is observed at degree 3. This is not only for this specific run of the training data split. The degree 3 gives the best results consistently. However, degrees from 3 to 6 have acceptable errors, and could be potentially used in estimating the function.

Dataset -
The given test set lacks noise and thus it has irreducible error very close to 0 (the total error graph is equal to the bias graph added to the variance graph). We can also say that data given is of degree 3 with respect to the input parameter, as the lowest error is obtained when our degree is 3, and the shape of the graph also indicates that Y = ax^3 + bx^2 + cx + d.