Task 2.1:

LinearRegression().fit()

=> This method is used to make our model adapt the given train data set. Like lets say we have a LinearRegression model of the type

$$F`(x) = cF(x)+d$$

What this method will do is it will try to find the values of the coefficients c and d such that the residual sum of squares between the observed targets in the dataset. This method will result in a trained model trained according to the dataset given to it. This function will also provide us the coefficients in case we are asked to report them.

Task 2.2.2:

DEG	BIAS	VARIANCE
1	825.7289831480639	47306.85757561664
2	812.4509076827252	67191.30137504089
3	109.10573402615323	67953.6748240348
4	121.44246221222923	71327.87728363136
5	113.07820970464937	101549.03033990711
6	123.84094901613075	108431.14726089456
7	118.62947964488384	124293.6425847867
8	125.3780016629125	136730.53241186385
9	135.94564814026506	161557.0915440239
10	151.0684104889412	157179.85718366923
11	129.5570173395351	268972.71168441855

12	132.0459148247733	235027.514907291
13	106.60823662353337	340587.0933027925
14	127.3413140889179	329287.0323314727
15	163.51567680553188	355464.8142334062
16	179.4687794104459	389598.1579159183
17	246.27371965647725	364509.10305524955
18	253.17429112229829	415356.76927538
19	325.01485692328225	374849.7515611527
20	329.74823606863873	435460.5708354182

Bias-Analysis: We find that the bias was initially very high uptil degree 2 as for these degrees our model is a victim of under-fitting. But from third degree onwards we experience a great fall in the values of the bias that is becuase the given test data resembles a degree 3 graph to a very high extent. The values remain approximately same uptil degree 13 and then they start to increase. One of the most reasonable reasons for this rise is the fact that as we increase our degree we basically increase the number of our local maximas and minimas so what happens is in order to have these number of maximas and minimas what happens is our model tries to adapt to the given values and at the same happens to achieve its required number of maximas and minimas which in turn result in a deviation from the actual values therefore the bias increase. So due to increase in model complexity beyond a certain value in order to maintain the complexity as well as the data set compatibility our model deviates from the values and which in turn is responsible for the increase in bias after 13^{th} degree (approx).

Variance-Analysis: We find that Variance remained almost similar from degree 2 to 4. After degree 4, the nature of Variance is

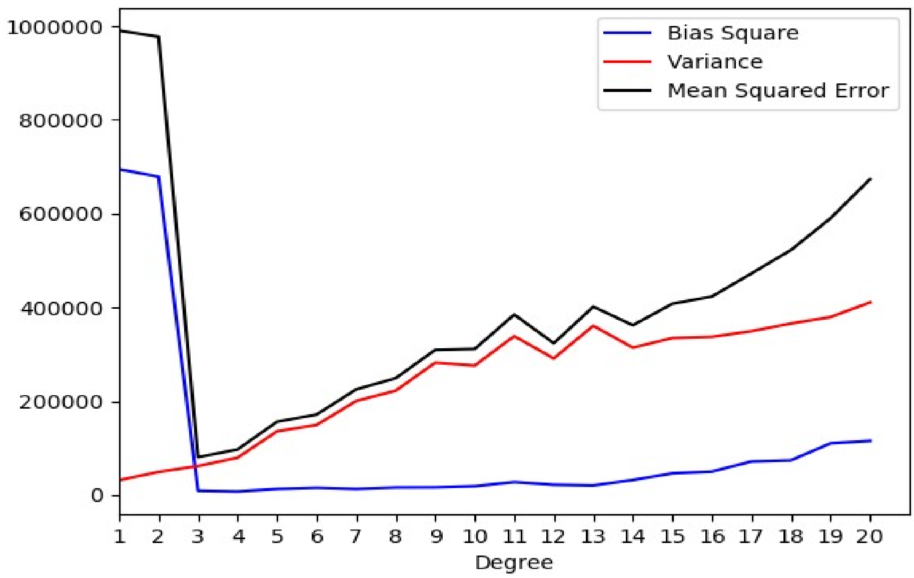
increasing. A model with High Variance is high in complexity. It gives high accuracy on the familiar training data. But when slightly different test data is presented to this model, its prediction varies considerably, which results in low accuracy. This is a sign of Overfitting. So, it's observed that Overfitting results from a Model that is high in complexity, i.e., High in Variance.
Deeper-Analysis: Our goal is to have a model that is low in Bias & low in Variance. To avoid High Bias, we will need to increase the complexity of our model & similarly, to prevent High Variance, we will need to decrease our model's complexity. We find the perfect balance at degree 3.

Task 2.3:

DEG	IRREDUCIBLE ERROR
1	-4.365575e-11
2	-1.309672e-10
3	1.455192e-11
4	2.910383e-11
5	2.910383e-11
6	-2.910383e-11
7	-5.820766e-11
8	0.000000e+00
9	-5.820766e-11
10	2.910383e-11
11	-5.820766e-11
12	0.000000e+00
13	1.164153e-10
14	1.164153e-10
15	-1.164153e-10
16	0.000000e+00
17	2.328306e-10
18	-1.164153e-10
19	0.000000e+00
20	-3.492460e-10

In the above table first of all i would like to say that the zeroes there aren’t actually zeroes, its just due to the fact the values were extremely small such that they were rounded off to 0 to show them. Moreover we can judge that the values are extremely small and almost constant that is of the order of $\sim 10^{-10}$. More over what we understood of the irreducible error is that it arises due to the some unknown features that we cannot record. So, no matter how perfect our model is, how well we train it there will be some error due to these unknown factors. And as these factors will always remain there and these factors are constant irrespective of the model we are talking about so the values of the irreducible error remain constant within some error of measurement. The noise among the data will also contribute to this irreducible error and moreover on shuffling and random sampling the average noise of the random sampling of data should remain constant so overall the irreducible error will remain constant. And the negative values correspond to those factors that we aren’t actually able to measure in life in form of mathematics so we get these absurd negative values.

Task 2.4:



From the Bias – Variance Tradeoff we know the optimal model is the one with degree 3. As for the optimality we try to minimize

variance, bias and the total error at the same time. And from the graph it is evident that the optimality will occur at 3. Before degree 3 we can observe that the bias square is quite high therefore bias is high that is the model is underfitting that is we are deviating from the actual values a lot. At the same time the variance is low as though the deviation from the actual values is high but the values output by the model with degree 1 and 2 are quite close to themselves that is the variation is quite low. This fact is also evident from the graph. But after 4^{th} degree on further increasing the complexity we find that both bias and variance begin to increase. This is due to the fact that now the model trained by us is now pretty more adaptive to the train data set. The model is able to perform quite well on the existing data but on the test data out model deviates a lot, this is a clear observation of overfitting. Our model has taken into account more number of features than it required leading to overfitting. Therefore the bias and variance keeps on increasing. We get a contradiction here which is that on increasing the complexity we should get a model with low bias as now our model is adapting more features of the given data but as explained in the ***bias-analysis*** of ***task 2.2.2*** ,we find that in order to maintain the number of required maximas and minimas in the graph of the corresponding complexity and at the same time adapt to the given train set our model finds a median of both which in turn happens to be a model of high bias.
On the type of data, we find that the data is quite consistent(in the starting values) but towards the larger values we have some outliers in the graph which makes the bias and variance values to increase on increase of the complexity of the model.
(y-axis represents error here)