Question - 1 : What is linearRegression.fit() function?

Answer: We have some training data with us (Single Variable or may be Multiple Variable) . Now we want to construct a linear function:- Y = summation(Ai.Xi)+C such that when we compare the values given by this constructed function with values given in training data set , then the mean square error must be minimized. linearRegression.fit() function in python does this task for us and selects values of all constants like Ai and C to minimize the mean square error.

Calculation of Bias and Variance :-

After dividing data set into 10 equally sized data sets , we calculated average bias , average variance and hence average irreducible error for each degree of polynomial (1-20). For a degree n and training model m, we first calculated bias for each point (x,y) given in test data set. So now we had 80 biases with us for a single training model, but we only wanted one for each training data set , so we take average of these 80 biases. Now we had to calculate final bias for a certain degree n but we had 10 biases with use because of 10 different functions construction (one for each training data set). So here again we calculated average of these 10 biases, finally leaving us with one average bias for each degree degree regression. Using similar method we calculate MSE and Variance for each degree polynomial regression. Using MSE , Variance and Bias we calculated Irrreducible Error.

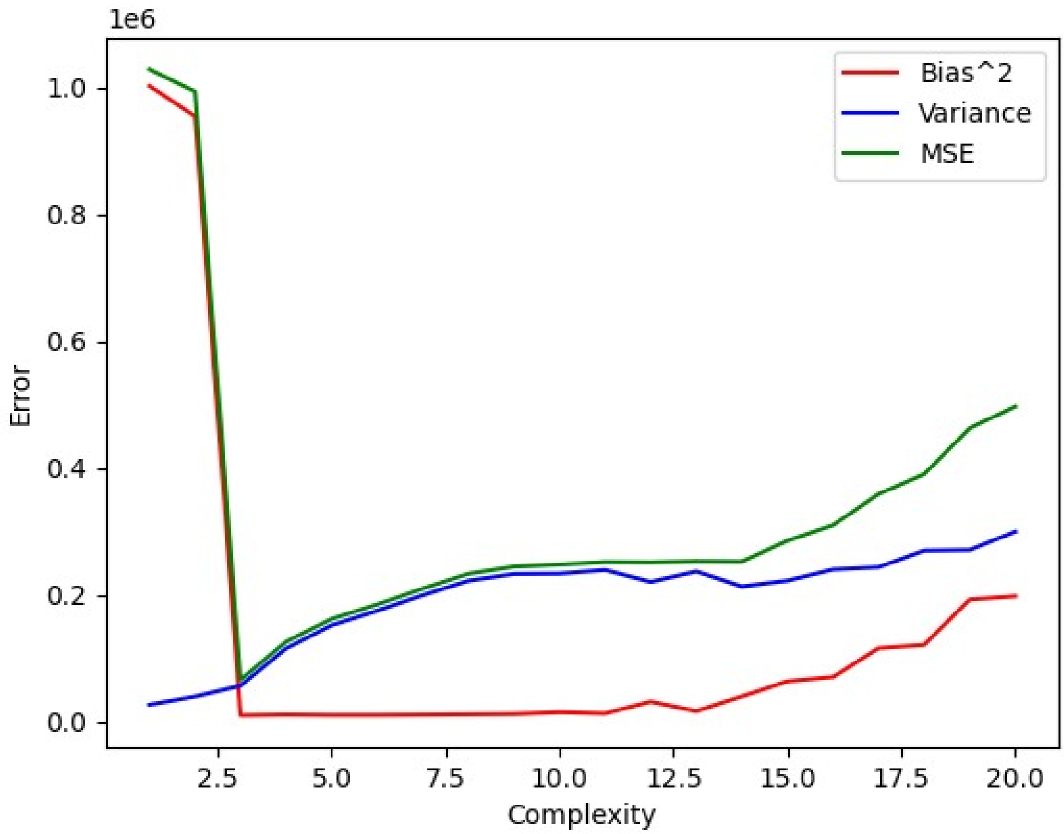
The data that we got after all the calculations is shown in the table below:-

Degree of Polynomial	Bias	Variance	Irreducible Error
1	819.7174357033888	25999.093009987868	-2.1827872842550278e-11
2	810.763391396119	39105.83381326804	3.055902197957039e-10
3	68.51085124220285	56095.89320972663	-8.731149137020111e-11
4	81.33971112338618	114907.29152950706	-1.1641532182693481e-10
5	78.95841418892078	151434.02790058558	-8.731149137020111e-10
6	78.36480094220735	174226.74500308032	1.7462298274040222e-10
7	86.91619731151918	198849.50274648867	-6.693881005048752e-10
8	90.32513553789045	221555.66219639863	2.3283064365386963e-10
9	92.35486444098777	232275.80526448222	-5.238689482212067e-10
10	97.94987666578288	232807.77103703268	0.0
11	91.1035858738162	238575.67800096897	2.3283064365386963e-10
12	125.84095426141899	219780.32854004213	6.984919309616089e-10
13	92.06347287289911	236241.2083400804	-2.9103830456733704e-10
14	130.17454852010573	212545.26284021224	9.022187441587448e-10
15	166.45932095455996	221715.2969090233	3.4924596548080444e-10
16	170.41744889300747	239357.88399177798	4.0745362639427185e-10
17	236.714748880766	242993.20230426235	-4.3655745685100555e-10
18	239.12519680886862	269052.2796714373	-2.9103830456733704e-10
19	304.8673610223364	270105.6024740014	2.3283064365386963e-10
20	305.44435432611056	299003.45980204793	5.820766091346741e-11

Few Remarks about our observations: -
Firstly the reason behind why we got negative irreducible error. The reason simply is because this error is of a order which is even smaller than the highest fraction accuracy that can be dealt by our machine. Also since each of the practically possible noises may not be mathematically representable , so what we mean is that negative errors coressponds to those noises which we can’t represent mathematically.

Polynomial of degree 3 seems to be most fitting for our testing data set (bias is least for this).

We will discuss other observations on the basis of graph shown below:-



We can clearly see in the graph that before degree=3 , variance is smaill but the bias is too high. This is what we call as underfitting data , as here the function is not too complex and does not fit our test data in them too well. At degree=3 , there is a sudden decline in value of bias and we can easily observe that bias is least at degree=3. After this the bias keeps on icreasing the maing it overfitting data. So for our model we believe that degree 3 polynomial is best suited to fit our traing and testing data sets. If we look at MSE and Variance , they keeps on increasing continuously . The reason is simple, as we make our function more complex , large variations will be induced while calculating predicted value at a certain x for different traing data sets. MSE will also increase because of more or less the same reason.