Task 2.1

Linear Regression fit explanation.

Consider an equation y = (b0)x + (b1). Lets load a dataset of x,y. and create a linear regression model using model = linearregression(). What model.fit() Does is it will calculate the optimal value of the b0, b1 using the existing input and output(x,y) as the arguments. We can say that its fits the model and also minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation.

And parameters used in this fit function are

1.fit_intercept(default is true): if the data is non-centered then you can ignore this is it is centered you can change it to 0.it is also means to calculate the intercept b1.

2. normalize(default is false): it means to normalize the data or not.

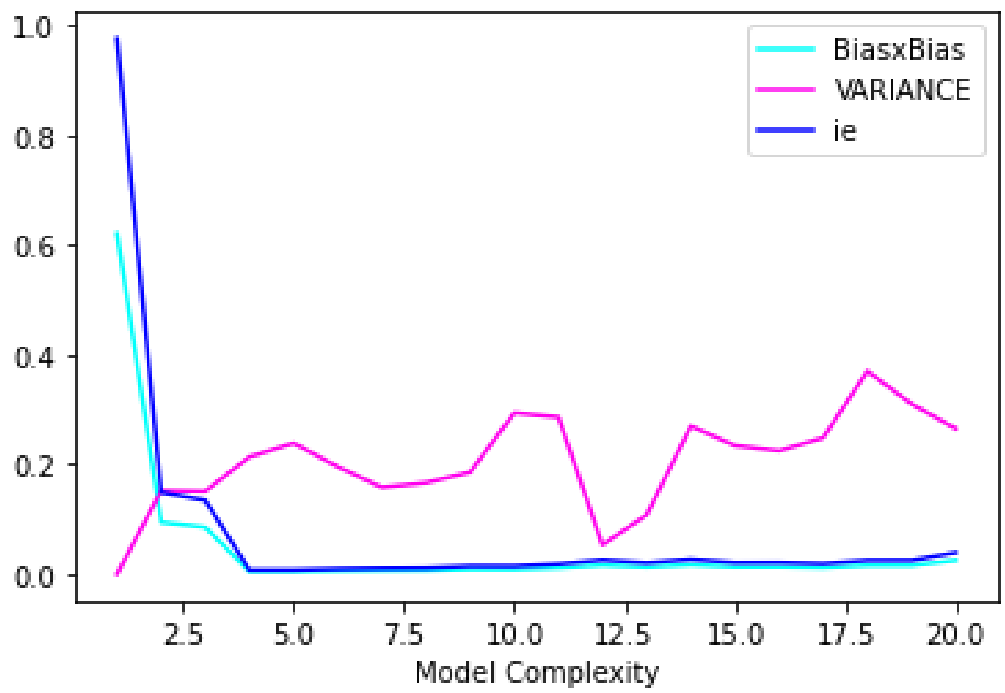
3.copy_X(default Is true): it means to copy (True)or overwrite(false) the input variables.

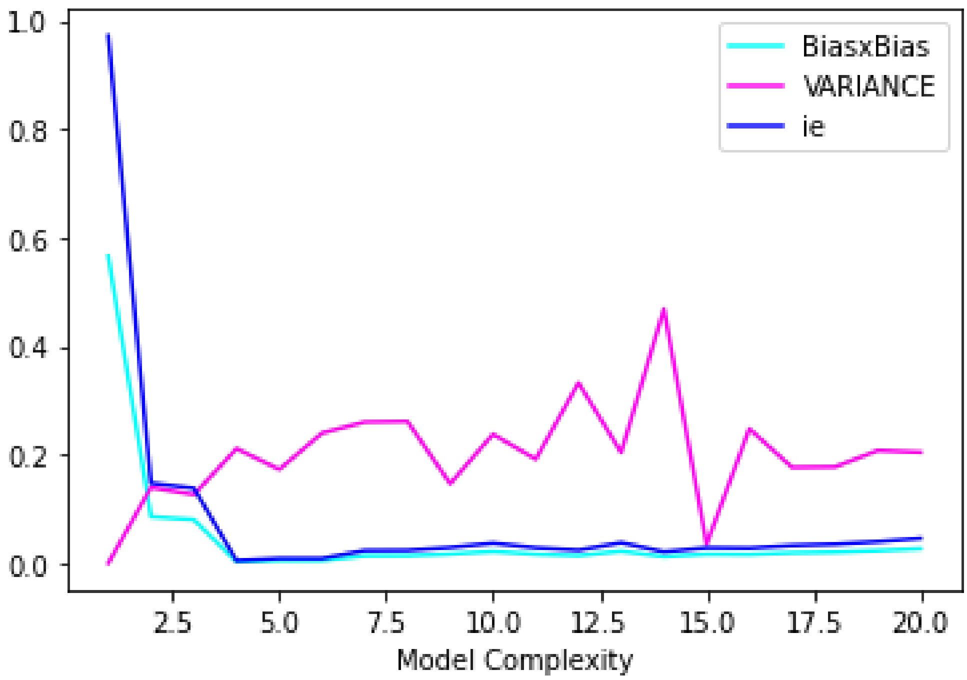
4.n_jobs(default is none): represents the number of jobs used in parallel computation. None usually means one job and -1 to use all processors.

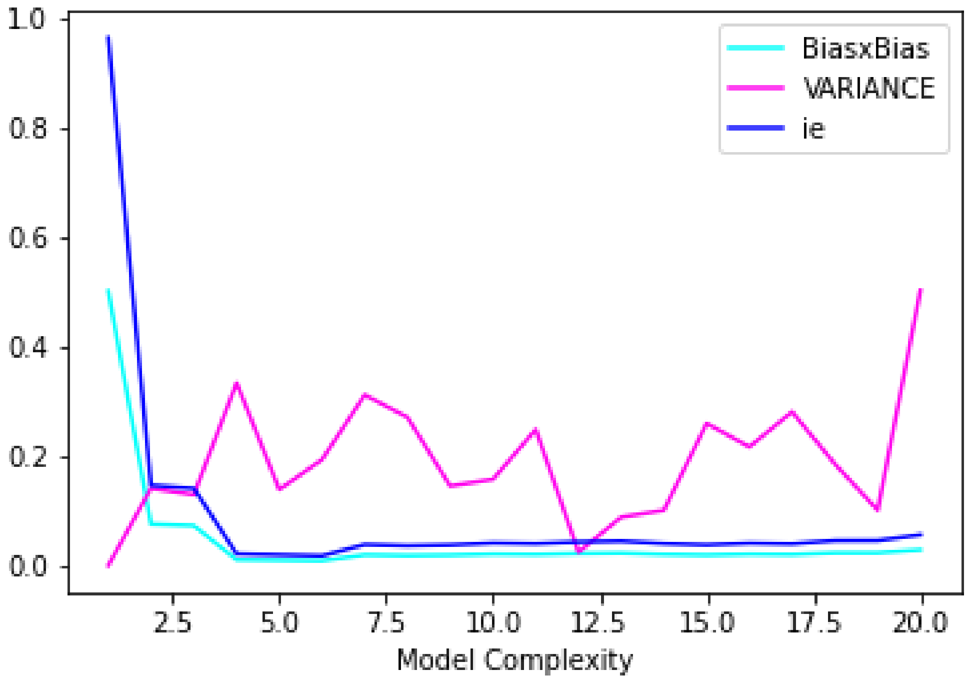
Task 2.2.2

Bias and variance report

Lets observe the image of the 1 to 20 degree polynomial bias and variance







The above three bias^2-variance graphs are for ten different random shuffled trained data sets.

Lets consider bias, as we can see bias is decreasing as the degree of the polynomial increases. Bias is defined as the error from test and training data.

	degree	bias	variance	ie
0	1	1.022138e+06	6.716188e-26	6.716188e-26
1	2	1.065244e+06	6.872619e-26	2.328306e-10
2	3	4.667749e+04	1.358261e-25	7.275958e-12
3	4	4.534952e+04	1.090225e-25	3.637979e-11
4	5	6.982530e+04	1.102916e-25	1.455192e-11
5	6	7.753558e+04	1.546451e-25	5.820766e-11
6	7	7.604292e+04	1.255059e-25	4.365575e-11
7	8	7.582988e+04	1.346137e-25	7.275958e-11
8	9	1.013257e+05	8.898099e-26	1.455192e-11
9	10	1.016502e+05	1.851387e-25	1.018634e-10
10	11	8.756666e+04	1.331225e-25	4.365575e-11
11	12	1.767134e+05	2.170787e-25	8.731149e-11
12	13	9.452490e+04	1.973566e-25	1.164153e-10
13	14	2.925688e+05	2.098976e-25	2.098976e-25
14	15	3.364573e+05	1.480080e-25	2.328306e-10
15	16	5.196326e+05	1.980154e-25	1.164153e-10
16	17	6.051125e+05	1.872566e-25	1.164153e-10
17	18	7.350124e+05	1.977378e-25	1.164153e-10
18	19	8.826569e+05	2.141082e-25	2.141082e-25
19	20	9.266564e+05	2.078983e-25	2.078983e-25

Variance is too low and is close to zero, so we normalized using linalg.norm in the code for the sake of comparisons of the graph. Lets consider the variance and its brief explanation Variance is the amount that the estimate of the target function will change if different training data was used. The target function is estimated from the training data by a machine learning algorithm, so we should expect the algorithm to have some variance. Here we can observe and say that as we change the train data sets variance is not changing much as there is not much of a difference while using the different pairs of trained data set.

Lets conclude the variation of bias and variance as the polynomial degree changes. Bias is decreasing as the polynomial degree decreasing and variance is slightly increasing which is a negligible and is very close to zero.

Task 3

Irreducible error(ie) table:

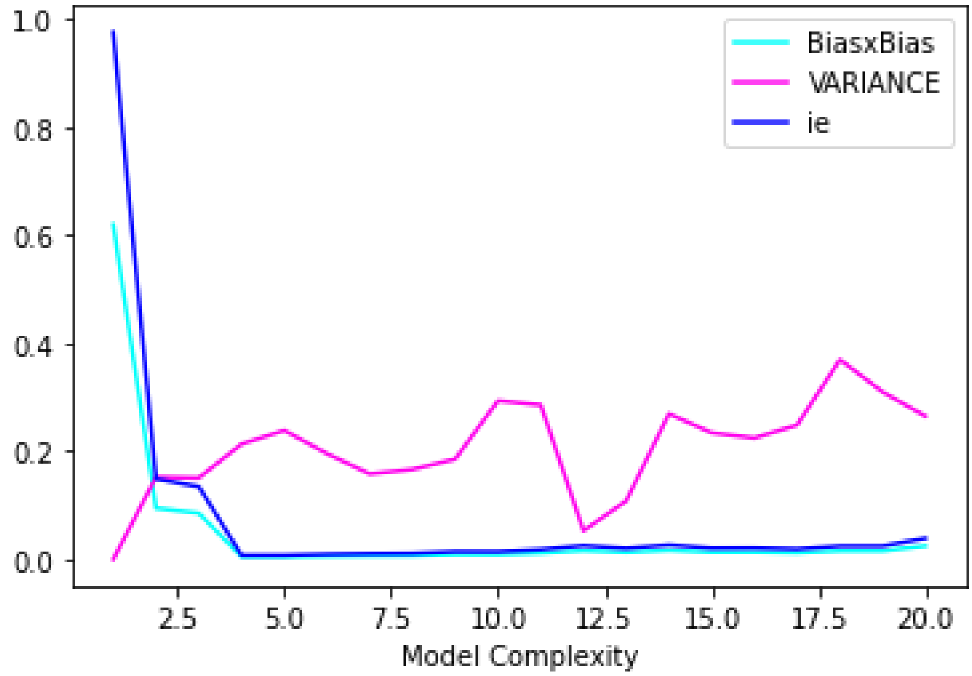
	degree	bias	variance	ie
0	1	1.022138e+06	6.716188e-26	6.716188e-26
1	2	1.065244e+06	6.872619e-26	2.328306e-10
2	3	4.667749e+04	1.358261e-25	7.275958e-12
3	4	4.534952e+04	1.090225e-25	3.637979e-11
4	5	6.982530e+04	1.102916e-25	1.455192e-11
5	6	7.753558e+04	1.546451e-25	5.820766e-11
6	7	7.604292e+04	1.255059e-25	4.365575e-11
7	8	7.582988e+04	1.346137e-25	7.275958e-11
8	9	1.013257e+05	8.898099e-26	1.455192e-11
9	10	1.016502e+05	1.851387e-25	1.018634e-10
10	11	8.756666e+04	1.331225e-25	4.365575e-11
11	12	1.767134e+05	2.170787e-25	8.731149e-11
12	13	9.452490e+04	1.973566e-25	1.164153e-10
13	14	2.925688e+05	2.098976e-25	2.098976e-25
14	15	3.364573e+05	1.480080e-25	2.328306e-10
15	16	5.196326e+05	1.980154e-25	1.164153e-10
16	17	6.051125e+05	1.872566e-25	1.164153e-10
17	18	7.350124e+05	1.977378e-25	1.164153e-10
18	19	8.826569e+05	2.141082e-25	2.141082e-25
19	20	9.266564e+05	2.078983e-25	2.078983e-25

As the basic definition of ie is mse-($bias^2$+variance). Here for our data sets provided variance is nearly negligible, and it is mainly dependent on bias and we can observe from the above graphs that the plot of $bias^2$ and ie are nearly merging with each other.

Considering variance is too negligible and bias is decreasing so the irreducible error is also decreasing.

Task 4

Lets take one of the $bias^2$, variance, ie graph.



As we know when the bias is too high we can say that it is underfitted and here we are ignoring variance because it is too low and it is close to zero. We can the model is underfitted upto 2.5. And it is nearly perfect in the ranges from 2.5 to 4. And if we consider the slight increment in the variance we can say it is overfitted as the polynomial degree increases to 20 or around the neighboring values of 20.

Here lets talk about the data given to us. The train data which we took ten equal parts and averaged to get ten different model sets to conduct test data to calculate the models performance. Here we can see the variance is too low so we can say that the model is performing too good with trained data sets.