MDL ASSIGNMENT 2

Name: Muskan Raina

Roll no.: 2021101066

TASK 1:

Linear regression is a technique used to model the relationship between a dependent variable say y, and an independent variable say x. We are to find the linear equation of the form $y = mx + b$ that predicts the values of y for given values of y .

$LinearRegression(\ ).fit(\ )$ is a method in the scikit-learn library that is used to train a linear regression model for a given training data. This function takes 2 input arrays X (2D array of independent variable x) and Y 1D array of dependent variable y).

During training, the function calculates the values of m and b that minimize the MSE (Mean Squared Error) which is the vertical distance between the Regression Line and each data point in the training dataset.

TASK 2:

We calculate the gradient of our cost function using gradient descent in order to reduce MSE. Gradient descent entails examining the error that our current weight produces, determining the gradient by utilising the cost function's derivative (i.e., the slope of the cost function using our current weight), and then adjusting our weight to travel in the gradient's opposite direction. Since the gradient indicates up the slope rather than down it, we must travel in the opposite direction of the gradient to reduce our inaccuracy.

Let us take our two parameters weight (w – independent) and bias b.

our cost function is taken to be the MSE

$$f(m,b) = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2$$

To derive the impact each parameter has on the final prediction, we use partial derivatives.

Differentiating wrt b,

$$\frac{df}{db} = \frac{1}{N}\sum_{i=1}^{n} 2\left(y_i - (mx_i + b)\right)\cdot(-1)$$

$$= \frac{1}{N}\sum_{i=1}^{n} 2\left((mx_i + b) - y_i\right)$$

Differentiating wrt m

$$\cancel{= \frac{1}{n}\sum_{i=1}^{n} 2\left(y_i - (mx_i + b)\right)^2}$$

$$\frac{df}{dm} = \frac{1}{N}\sum_{i=1}^{n} 2\left(y_i - (mx_i + b)\right)\cdot(-x_i)$$

$$= \frac{1}{N}\sum_{i=1}^{n} (-2x_i)\left(y_i - (mx_i + b)\right)$$

We use the partial derivatives to update the values of the parameters b and w.

<u>ALL OBSERVATIONS SHOWN WITH RESPECT TO SEED (100)</u>

<u>TASK 3:</u>

Observing the tabulated data, we see that as the degree of the polynomial function increases, the bias decreases and the variance increases.

Our computed bias starts out higher and the accompanying variance is relatively small. The bias starts to decline, and the variance rises as the degree increases.

The tabulated values are displayed below:

Bias	Variance
0.269347434813731	0.010543255949895651
0.08597088969438774	0.0012204083474801422
0.03330844622207401	0.0003701877481543818
0.025331018318202905	0.00044634557793805875
0.025252900944629356	0.0005465717386827752
0.026107711686110757	0.000744096814052807
0.025190983929929342	0.0008143086391643617
0.025362653957637992	0.0011837949232124478
0.029336508599132036	0.011992898866267112
0.03695491258452591	0.07380265782404047
0.04167806102583516	0.14941124918504756
0.04334779466550367	0.2067322478822584
0.04184663930937454	0.5232691189784373
0.10168716728642574	1.4521304710731764
0.17188469094542053	11.875190608071728

TASK 4:

The irreducible error is the error in the data that cannot be eliminated even with the best possible model. It is caused by the presence of noise in the data. Noise is the unwanted distortion in data.

From our calculations we see that the irreducible error is very small (of the order of 10^{-19} to 10^{-15}). These values denote that there is no significant noise in the dataset. The values are similar for all the models as the noise in the dataset remains the same across the different models as noise is a property of the given dataset and not the model. We notice that we obtain a few zeros and negative values as well. This indicates that some of the errors may be due to rounding errors caused by Python, rather than the noise in the dataset.

The tabulated values are displayed below:

```
Irreducible Error
------------------------
-1.734723475976807e-18
2.3852447794681098e-18
1.4094628242311558e-18
1.4094628242311558e-18
3.2526065174565133e-19
4.336808689942018e-19
1.0842021724855044e-19
6.505213034913027e-19
1.734723475976807e-18
0.0
5.551115123125783e-17
0.0
1.1102230246251565e-16
-2.220446049250313e-16
1.7763568394002505e-15
```

TASK 5:

Bias is the difference between the average prediction of our model and the correct value that we are trying to predict, and variance is how much the predictions for a given point varies between different models.

We observe the graph to see that initially our calculated bias has a higher value, and the corresponding variance is quite low. This indicates an underfitted model which means that the model does not perform well on training sets and thus will not perform well on test sets either.

As the degree increases, the bias begins to decrease, and the variance increases. This indicates that our model is now overfitting as the bias is low and the variance high. Overfitting models tend to extract too much information from the training set which leads to the model performing well on the training set but not on the test set.

The Optimum Model Complexity is the Model Complexity where the Mean Squared Error is minimized. In our graph we notice that the Optimum Model Complexity is achieved when the MSE is minimum, i.e., when the degree of the polynomial is 4.

TABLES AND GRAPHS:

The table of the calculated Degree, Bias, Bias Square, Variance, MSE and Irreducible Error:

Degree	Bias	Bias Square	Variance	MSE	Irreducible Error
1	0.269347434813731	0.1145311821094268	0.010543255949895651	0.12507443805932245	-1.734723475976807e-18
2	0.08597088969438774	0.012174557431817525	0.0012204083474801422	0.01339496577929767	2.3852447794681098e-18
3	0.03330844622207401	0.004720843707298192	0.0003701877481543818	0.005091031455452575	1.4094628242311558e-18
4	0.025331018318202905	0.0042506178758128	0.00044634557793805875	0.00469696345375086	1.4094628242311558e-18
5	0.025252900944629356	0.004243467433071845	0.0005465717386827752	0.00479003917175462	3.2526065174565133e-19
6	0.026107711686110757	0.004308695975445136	0.000744096814052807	0.005052792789497943	4.336808689942018e-19
7	0.025190983929929342	0.004270334590114568	0.0008143086391643617	0.00508464322927893	1.0842021724855044e-19
8	0.025362653957637992	0.0043222678329505005	0.0011837949232124478	0.005506062756162949	6.505213034913027e-19
9	0.029336508599132036	0.004605441853365649	0.011992898866267112	0.016598340719632763	1.734723475976807e-18
10	0.03695491258452591	0.00810039453234688	0.07380265782404047	0.08190305235638735	0.0
11	0.04167806102583516	0.0115330656455911	0.14941124918504756	0.1609443148306387	5.551115123125783e-17
12	0.04334779466550367	0.010808531481865084	0.2067322478822584	0.2175407793641235	0.0
13	0.04184663930937454	0.006667950424241846	0.5232691189784373	0.5299370694026793	1.1102230246251565e-16
14	0.10168716728642574	0.09336099297354818	1.4521304710731764	1.5454914640467243	-2.220446049250313e-16
15	0.17188469094542053	0.508886698151225	11.875190608071728	12.384077306222954	1.7763568394002505e-15

The graph visualizing the Degree, Bias Squared, Variance, MSE and Irreducible Error: