Assignment-2 report

3. The obtained values indicate the average bias and variance for different polynomial degrees in your model. Let's analyze the results and discuss how bias and variance change as you vary your function classes (polynomial degrees):
Observations:
Degree 1:

Average Bias: 1.0305
Average Variance: 0.0367
Observation: The model with a linear function (degree 1) has a relatively high average bias, suggesting that it might be too simple to capture the underlying patterns in the data. The low variance indicates that the model's predictions do not fluctuate significantly across different subsets of the training data.

Degree 2:

Average Bias: 1.0055
Average Variance: 0.0659
Observation: The bias slightly decreases, indicating that the quadratic function (degree 2) captures more complexity. However, the variance increases, suggesting that the model becomes more sensitive to the training data.

Degree 3:

Average Bias: 0.0299
Average Variance: 0.0864
Observation: The bias significantly decreases, indicating that the cubic function (degree 3) fits the data well. The variance remains relatively stable, suggesting that the model generalizes consistently across different training subsets.

Degree 4 - Degree 6:

Observation: The bias continues to decrease, indicating that the models fit the data more accurately. The variance, however, starts to increase, suggesting that the models become more sensitive to variations in the training data.

Degree 7 - Degree 10:

Observation: The bias remains relatively low, indicating that the high-degree polynomials can fit the data well. However, the variance increases, suggesting that the models become increasingly sensitive to variations in the training data, potentially leading to overfitting.

Overall Trend:

As the polynomial degree increases, the models become more flexible and can fit the training data more closely.

Initially, the bias decreases as the models become more complex, but eventually, the variance starts to increase, indicating a trade-off between bias and variance.

The optimal degree might be one that achieves a good balance between bias and variance, avoiding underfitting and overfitting.

4. The consistently small and close-to-zero values of the irreducible error across different polynomial degrees suggest that, in the context of this specific analysis, the irreducible error does not significantly change as the complexity of the class function (polynomial degree) varies.

The irreducible error represents the inherent noise or randomness in the data that cannot be captured by any model. It is expected to be a constant, fundamental component of the mean squared error (MSE) that persists regardless of the model's complexity. In an ideal scenario, the irreducible error should remain unchanged as you vary the class function.

The small variations observed in the irreducible error values can be attributed to numerical precision and rounding errors inherent in floating-point arithmetic during the computations. In practical terms, these small variations are negligible and do not impact the overall interpretation of the irreducible error.

In conclusion, the consistent and small values of irreducible error across different polynomial degrees indicate that the fundamental noise level in the data remains relatively stable, and the variations observed can be attributed to numerical artifacts rather than meaningful changes in the inherent noise as the class function varies.

TABLE FOR TASKS 3 AND 4:

Bias_squared	Variance	Irreducible Error	MSE
1.0305	0.0367	4.7878e-16	1.0673
1.0055	0.0659	7.9103e-16	1.0714
0.0299	0.0864	1.9429e-16	0.1163
0.0081	0.1004	0.0	0.1085
0.0483	0.0876	1.8041e-16	0.1359
0.0276	0.128	5.5511e-17	0.1557
0.0226	0.1074	5.5511e-17	0.1299
0.0327	0.1491	-1.6653e-16	0.1817
0.1267	0.1382	1.1102e-16	0.2649
0.2715	0.1524	1.6653e-16	0.4240

5. Observations:

Underfitting (Degree 1-2):
Bias² is relatively high, indicating that the model is too simple to capture the underlying patterns in the data.
Variance is relatively low.
Total error is dominated by Bias².

Balanced Model (Degree 3-6):
Bias² decreases, indicating that the model fits the data better.
Variance starts to increase, suggesting a balance between fitting the data and generalizing.
Total error decreases.

Overfitting (Degree 7-10):
Bias² remains low.
Variance increases, indicating that the model is becoming too sensitive to variations in the training data.
Total error starts to increase due to high variance.

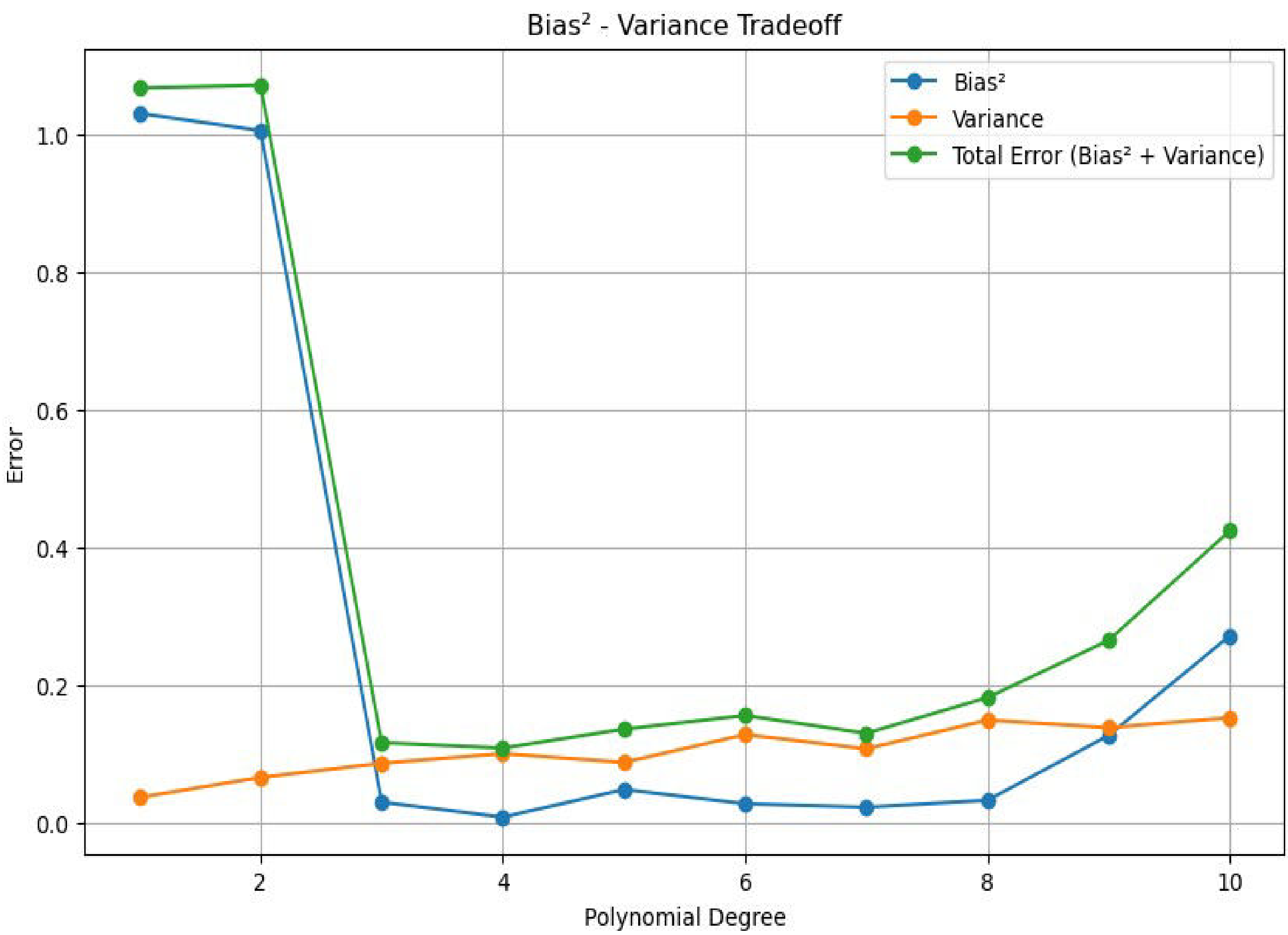
Overall Trend:

Initially, Bias² dominates, indicating underfitting.
There is a sweet spot (around Degree 4-6) where Bias² and Variance are balanced.

Beyond that point, increasing model complexity leads to overfitting, as Variance starts to dominate.

Type of Data:

The analysis suggests that the data might have a moderately complex underlying structure, and a polynomial degree around 3-5 achieves a good balance between fitting the data and avoiding overfitting.



ANSWER 1-

(a) We will define a linear relationship between dependent variable Y and independent variable X as follows:

$$Y=mX+c$$

The loss is the error in our predicted value of **m** and **c**. Our goal is to minimize this error to obtain the most accurate value of **m** and **c**.
We will use the Mean Squared Error function to calculate the loss. There are three steps in this function:

1. Find the difference between the actual y and predicted y value($y = mx + c$), for a given x.

2. Square this difference.

3. Find the mean of the squares for every value in X.

$$E=1/n\ \Sigma(y_i - \bar{y}_i)^2$$

Here y_i is the actual value and $\bar{y}_i$ is the predicted value. Lets substitute the value of $\bar{y}_i$:

$$E= 1/n\ \Sigma(y_i - (mx_i + c))^2$$

The Gradient Descent Algorithm

Gradient descent is an iterative optimization algorithm to find the minimum of a function. Here that function is our Loss Function.

1. Initially let $m = 0$ and $c = 0$. Let L be our learning rate. This controls how much the value of **m** changes with each step. L could be a small value like 0.0001 for good accuracy.

2. Calculate the partial derivative of the loss function with respect to m, and plug in the current values of x, y, m and c in it to obtain the derivative value **D**.

$$D_m = \frac{1}{n}\sum_{i=0}^{n} 2(y_i - (mx_i + c))(-x_i)$$

$$D_m = \frac{-2}{n}\sum_{i=0}^{n} x_i(y_i - \bar{y}_i)$$

D_m is the value of the partial derivative with respect to **m**. Similarly lets find the partial derivative with respect to **c**, Dc :

$$D_c = \frac{-2}{n}\sum_{i=0}^{n} (y_i - \bar{y}_i)$$

3. Now we update the current value of **m** and **c** using the following equation:

$$m = m - L \times D_m$$

$$c = c - L \times D_c$$

4. We repeat this process until our loss function is a very small value or ideally 0 (which means 0 error or 100% accuracy). The value of **m** and **c** that we are left with now will be the optimum values.

(b) Our equation will now look like this:

Y= mX+nZ+c

Procedure is exactly same except for the fact that we have to now find derivative with respect to each variable, i.e., we will now find with respect to m,c and n. We will also find D_n and will update n accordingly by the same formula n=n- L X $D_{n.}$ We will repeat until n,m and c become very small values.

ANSWER-2-

We have been given 6 data points.We will verify the formula MSE=bias^2 +variance for each data point.

The three models given to us are:

$$f_1 (x) = 2x^2 + 3x + 1$$
$$f_2 (x) = x^2 + 3x$$
$$f_3 (x) = 2x^2 + 2x + 1$$

We will use the following thee formulas:

1. Bias=$E_i [\hat{f}_i (x)] − f (x)$
2. Variance = $E_i [(\hat{f}_i (x) − E_i [\hat{f}_i(x)])^2]$
3. MSE= $E_i [(f (x) − f\hat{}(x))^2]$

where f (x) represents the true value, $\hat{f}i$ (x) represents the value predicted by the ith model, Ei [·] is the expectation over all models.

1. For the data point (-2,5):

$$f_1(-2)= 2(-2)^2 + 3(-2) +1=3$$
$$f_2(-2)= (-2)^2 + 3(-2)= -2$$
$$f_3(-2)= 2(-2)^2 + 2(-2) + 1=5$$

$E_i [\hat{f}_i (x)] = (3-2+5)/3 = 2$

Bias= 2- y = 2-5 = -3

Now we will calculate variance.

$(f_1(-2) - E_i [\hat{f}_i (x)])^2 = (3 – 2)^2 = 1$

$(f_2(-2) - E_i [\hat{f}_i (x)])^2 = (-2 – 2)^2 = 16$

$(f_3(-2) - E_i [\hat{f}_i (x)])^2 = (5 – 2)^2 = 9$

Variance=(1+ 16+ 9)/3= 26/3

Now we will calculate MSE.

$(y- f_1(-2))^2= (5 – 3)^2= 4$

$(y- f_2(-2))^2= (5 + 2)^2= 49$

$(y- f_3(-2))^2= (5 – 5)^2= 0$

MSE = (4 + 49+0)/3 = 53/3

$(-3)^2 + 26/3= 9+ 26/3 = 53/3$

Hence, the formula $Bias^2$+ Variance= MSE is true.

2. For the data point (-1,0):

$$f_1(-1)= 2(-1)^2 + 3(-1) +1=0$$
$$f_2(-1)= (-1)^2 + 3(-1)= -2$$
$$f_3(-1)= 2(-1)^2 + 2(-1) + 1=1$$

$E_i [\hat{f}_i (x)] = (0-2+1)/3 = -1/3$

Bias= -1/3 - y = -1/3 – 0 = - 1/3

Now we will calculate variance.

$(f_1(-1) - E_i [\hat{f}_i (x)])^2 = (0 + 1/3)^2 = 1/9$

$(f_2(-1) - E_i [\hat{f}_i (x)])^2 = (-2 + 1/3)^2 = 25/9$

$(f_3(-1) - E_i [\hat{f}_i (x)])^2 = (1 + 1/3)^2 = 16/9$

Variance=(1/9+ 16/9+ 25/9)/3= 14/9

Now we will calculate MSE.

$(y- f_1(-1))^2= (0 – 0)^2= 0$

$(y- f_2(-1))^2= (0 + 2)^2= 4$

$(y- f_3(-1))^2= (0 – 1)^2= 1$

MSE = (4 + 1+0)/3 = 5/3

$(-1/3)^2 + 14/9= 1/9+ 14/9 = 5/3$

Hence, the formula $Bias^2$+ Variance= MSE is true.

3. For the data point (0,1):

$$f_1(0)= 2(0)^2 + 3(0) +1=1$$
$$f_2(0)= (0)^2 + 3(0)= 0$$
$$f_3(0)= 2(0)^2 + 2(0) + 1=1$$

$E_i [\hat{f}_i (x)] = (1+0+1)/3 = 2/3$

Bias= 2/3- y = 2/3-1 = -1/3

Now we will calculate variance.

$(f_1(0) - E_i [\hat{f}_i (x)])^2 = (1 – 2/3)^2 = 1/9$

$(f_2(0) - E_i [\hat{f}_i (x)])^2 = (0 – 2/3)^2 = 4/9$

$(f_3(0) - E_i [\hat{f}_i (x)])^2 = (1 – 2/3)^2 = 1/9$

Variance=(1/9+ 4/9+ 1/9)/3= 2/9

Now we will calculate MSE.

$(y- f_1(0))^2= (1 – 1)^2= 0$

$(y- f_2(0))^2= (1 - 0)^2= 1$

$(y- f_3(0))^2= (1 – 1)^2= 0$

$$MSE = (0+ 1+0)/3 = 1/3$$

$$(-1/3)^2 + 2/9= 1/9+ 2/9 = 1/3$$

Hence, the formula $Bias^2$+ Variance= MSE is true.

4. For the data point (1,4):

$$f_1(1)= 2(1)^2 + 3(1) +1=6$$
$$f_2(1)= (1)^2 + 3(1)= 4$$
$$f_3(1)= 2(1)^2 + 2(1) + 1=5$$
$$E_i [\hat{f}_i (x)] = (6+5+4)/3 = 5$$
$$Bias= 5 - y = 5 – 4 = 1$$

Now we will calculate variance.

$$(f_1(1) - E_i [\hat{f}_i (x)])^2 = (6 - 5)^2 = 1$$
$$(f_2(1) - E_i [\hat{f}_i (x)])^2 = (4 - 5)^2 = 1$$
$$(f_3(1) - E_i [\hat{f}_i (x)])^2 = (5 - 5)^2 = 0$$
$$Variance=(1+1+0)/3= 2/3$$

Now we will calculate MSE.

$$(y- f_1(1))^2= (4 – 6)^2= 4$$
$$(y- f_2(1))^2= (4 - 4)^2= 0$$
$$(y- f_3(1))^2= (4 – 5)^2= 1$$
$$MSE = (4 + 1+0)/3 = 5/3$$

$$(1)^2 + 2/3= 1+ 2/3 = 5/3$$

Hence, the formula $Bias^2$+ Variance= MSE is true.

5. For the data point (2,11):

$$f_1(2)= 2(2)^2 + 3(2) +1=15$$
$$f_2(2)= (2)^2 + 3(2)= 10$$
$$f_3(2)= 2(2)^2 + 2(2) + 1=13$$
$$E_i [\hat{f}_i (x)] = (10+15+13)/3 = 38/3$$
$$Bias= 38/3 - y = 38/3 – 11 = 5/3$$

Now we will calculate variance.

$$(f_1(2) - E_i [\hat{f}_i (x)])^2 = (15- 38/3)^2 = 49/9$$
$$(f_2(2) - E_i [\hat{f}_i (x)])^2 = (10 - 38/3)^2 = 64/9$$
$$(f_3(2) - E_i [\hat{f}_i (x)])^2 = (13 - 38/3)^2 = 1/9$$
$$Variance=(49/9+ 64/9+ 1/9)/3= 38/9$$

Now we will calculate MSE.

$$(y - f_1(2))^2 = (11 - 15)^2 = 16$$
$$(y - f_2(2))^2 = (11 - 10)^2 = 1$$
$$(y - f_3(2))^2 = (11 - 13)^2 = 4$$
$$MSE = (4 + 1 + 16)/3 = 7$$

$$(5/3)^2 + 38/9 = 25/9 + 38/9 = 7$$

Hence, the formula $Bias^2 + Variance = MSE$ is true.

6. For the data point (3,22):

$$f_1(3) = 2(3)^2 + 3(3) + 1 = 28$$
$$f_2(3) = (3)^2 + 3(3) = 18$$
$$f_3(3) = 2(3)^2 + 2(3) + 1 = 25$$
$$E_i[\hat{f}_i(x)] = (28+18+25)/3 = 71/3$$
$$Bias = 71/3 - y = 71/3 - 22 = 5/3$$

Now we will calculate variance.

$$(f_1(3) - E_i[\hat{f}_i(x)])^2 = (28 - 71/3)^2 = 169/9$$
$$(f_2(3) - E_i[\hat{f}_i(x)])^2 = (18 - 71/3)^2 = 289/9$$
$$(f_3(3) - E_i[\hat{f}_i(x)])^2 = (25 - 71/3)^2 = 16/9$$
$$Variance = (169/9 + 289/9 + 16/9)/3 = 158/9$$

Now we will calculate MSE.

$$(y - f_1(3))^2 = (22 - 28)^2 = 36$$
$$(y - f_2(3))^2 = (22 - 18)^2 = 16$$
$$(y - f_3(3))^2 = (22 - 25)^2 = 9$$
$$MSE = (36 + 16 + 9)/3 = 61/3$$

$$(5/3)^2 + 158/9 = 25/9 + 158/9 = 61/3$$

Hence, the formula $Bias^2 + Variance = MSE$ is true.