Machine, Data and Learning: Assignment 1

Shubhangi Dutta (2018113004), Rishav Kundu (2019121007)

February 14, 2020

1 Bias and Variance

The formula for bias is given as

$$Bias = (E[\hat{f}(x)] - f(x))$$

Therefore,

$$Bias^2 = (E[\hat{f}(x)] - f(x))^2$$

The formula for variance is given as

$$Variance = E[\hat{f}(x) - E[\hat{f}(x)])^2]$$

The above formulas are defined for a single value of x. Hence, the expectation $E[\hat{f}(x)]$ is defined over the different values of $\hat{f}(x)$ obtained by training the same model on different training set partitions.

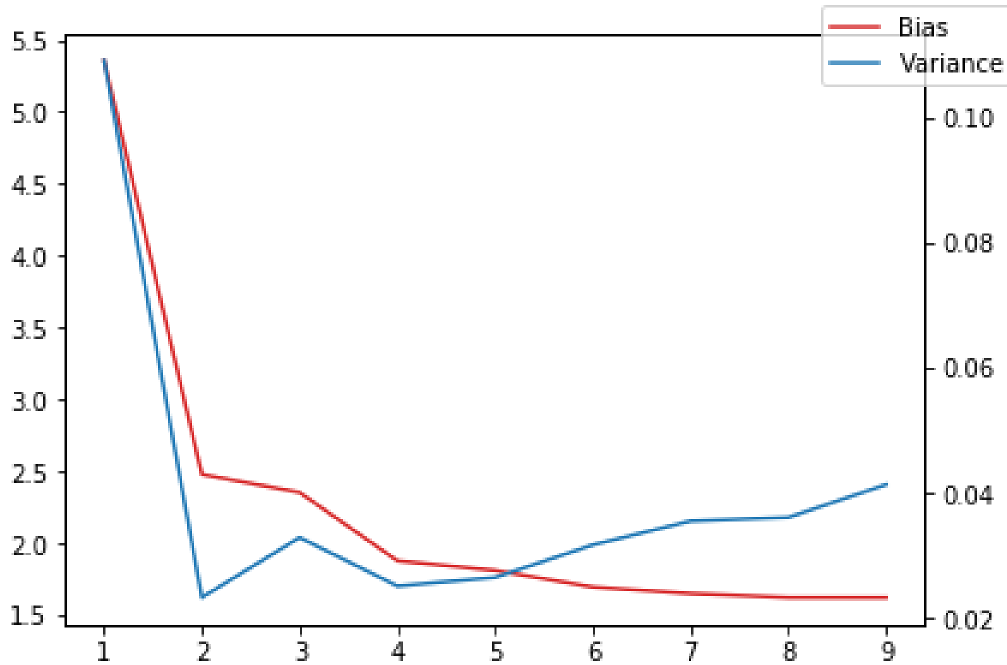
2 Q1

2.1 Bias and Variance table for Polynomial Fit

Degree	Bias	Variance
1	30.37408273224508	0.09341804232251934
2	6.36517276002383	0.02342602657696328
3	5.2694201717911	0.032432266902508056
4	3.1505752052574882	0.024700107894447286
5	2.929418757965218	0.026853397616000967
6	2.5924918432950887	0.025693386573740703
7	2.4297943137364473	0.03185979691247053
8	2.401411060979298	0.04025175952926954
9	2.4046636467383133	0.042469600227830584
10	2.417341604358687	0.047025847634951844
11	2.2129311318120677	0.039547099980470186
12	2.212837336615444	0.040014205734798594
13	2.04333525171306	0.04465927459234133
14	2.044422212520452	0.054205383136040355
15	2.048575476047638	0.05184750688113316
16	2.055511225205248	0.05780882443374418
17	1.5676562220889405	0.041844397410982295
18	1.580772918328726	0.04997588536630306
19	1.5753515315568343	0.04587184307778622
20	1.5700026631879862	0.05801091925152769
21	1.0749491166688236	0.09679378957868845
22	1.0773198062251255	0.11390297913298957
23	1.0665098885278717	0.11908804588652822
24	0.9482893416038309	0.06419302845892338
25	0.8195925886965155	0.04129120796111043
26	0.48111794650702805	0.14104241519819197
27	0.5641635223612983	0.12850830831675497
28	0.6162690155838716	0.11799193084262229
29	0.15763860143872435	0.11608161243502352
30	0.136425684845244	0.04422191964882733

2.2 Bias-Variance Vs Degree Plot for Polynomials Degree 1-9

2.2.1 Plot

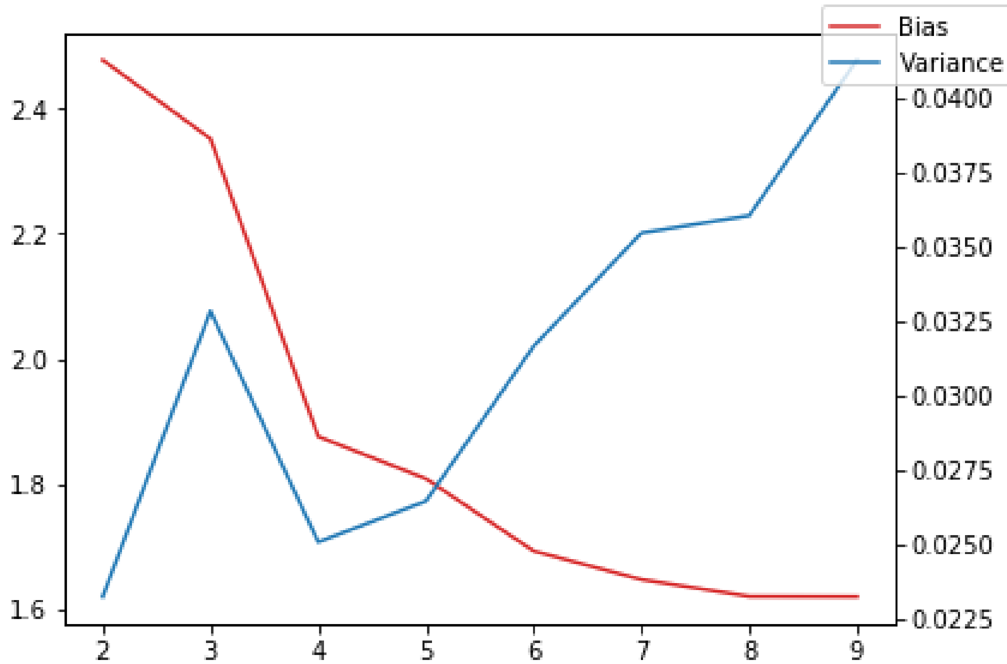


2.2.2 Observations and Analysis

- Both bias and variance are very large for the linear fit. The model is underfit and performs poorly for the given data.
- For higher degrees (>1, < 6), both bias and variance decrease. The models are better fit to the data points, so both decrease.
- Variance increases again for even higher degree polynomials (>6) but it is not significant compared to the drop between degree 1 and 2. The model is overfitting, and hence is not a good fit for the test data and is accounting for small deviations, making variance high.

2.3 Bias-Variance Vs Degree Plot for Polynomials Degree 2-9

We remove the first point to better view the changes between the other degrees.

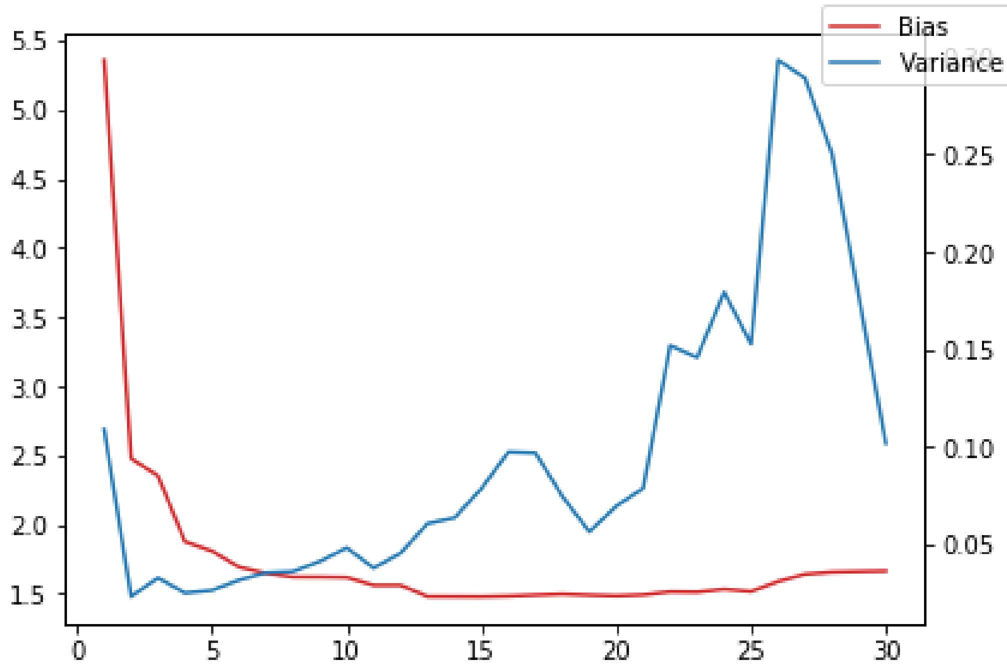


2.3.1 Observations and Analysis

- Bias is high but variance is low for the degree 2 fit. The model is underfit and performs poorly for the test data.
- For higher degrees (> 2, < 5), both bias and variance decrease. The models are better fit to the data points, so both decrease.
- The intersection point is at the 5 degree polynomial. Both bias and variance are low and this is optimal fit.
- Variance increases again for even higher degree polynomials (>6) and it is significant compared to the drop in bias. The model is overfitting, and hence is not a good fit for the test data and is accounting for small deviations, making variance high.

2.4 Bias-Variance Vs Degree Plot for Polynomials Degree 1-30

We increase the degree of the polynomial to get a wider view of the data.



2.4.1 Observations and Analysis

- Both bias and variance are very large for the linear fit. The model is underfit and performs poorly for the given data.
- For higher degrees (> 1, < 10), both bias and variance decrease. The models are better fit to the data points, so both decrease.
- The intersection point is at the 5 degree polynomial. Both bias and variance are low and this is optimal fit.
- Variance increases again sharply for even higher degree polynomials (¿10). Bias keeps reducing in approximately linear way. The model is overfitting, and hence is not a good fit for the test data and is accounting for small deviations, making variance high.
- In some cases, such as at degree 24-25 and again at 30. In these cases, the model incidentally fit well, probably due to the nature of the data.

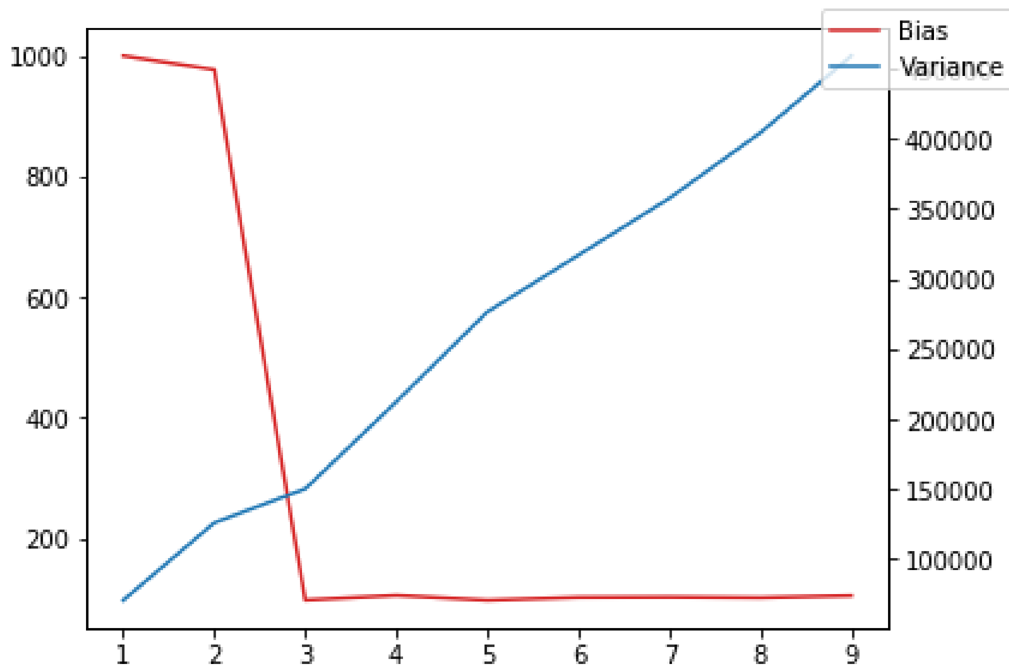
3 Q2

3.1 Bias and Variance table for Polynomial Fit

Degree	Bias	Variance
1	999.6141239858127	70545.48914575046
2	977.0461983931082	125870.85554877334
3	96.90061979570203	150073.7395464768
4	104.43825033995597	212235.70832526154
5	96.63950688681115	276388.48025474057
6	101.23529987681114	316863.49843748985
7	101.66255879943755	357510.98475735466
8	100.74432611287484	404286.670685786
9	103.99753380044281	459132.37837248633

3.2 Bias-Variance Vs Degree Plot for Polynomials Degree 1-9

3.2.1 Plot



3.2.2 Observations and Analysis

- Both bias and variance are very large for the linear fit. The model is underfit and performs poorly for the given data.

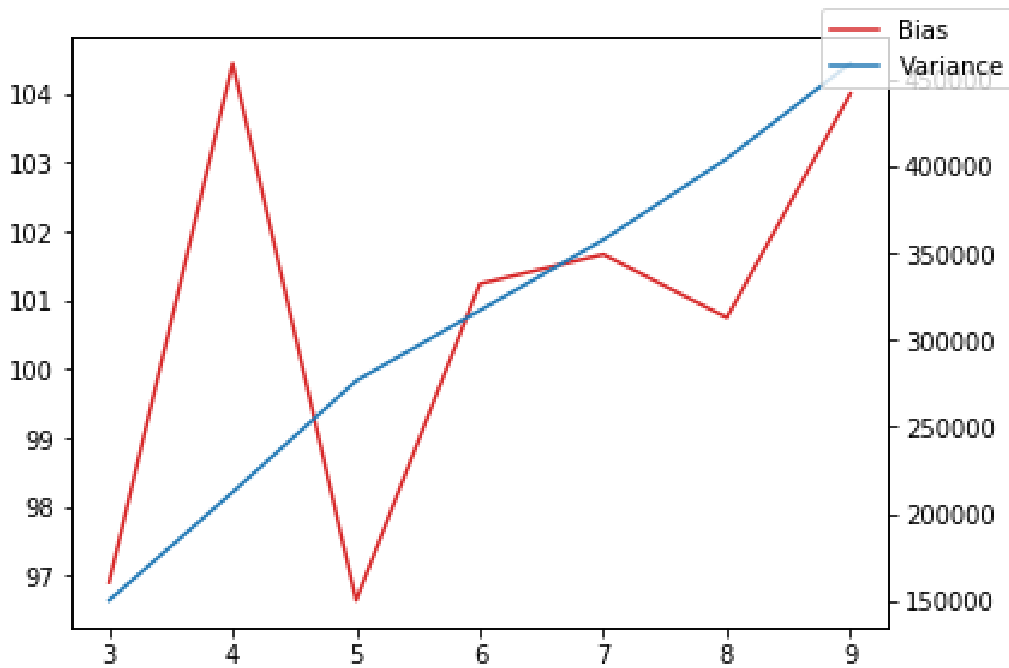
- For higher degrees (>1, <3), both bias and variance decrease. The models are better fit to the data points, so both decrease.

- Variance increases again for even higher degree polynomials (>3) but it is not significant compared to the drop between degree 1 and 2. The model is overfitting, and hence is not a good fit for the test data and is accounting for small deviations, making variance high.

3.3 Bias-Variance Vs Degree Plot for Polynomials Degree 3-9

We remove the first point to better view the changes between the other degrees.

3.3.1 Plot



3.3.2 Observations and Analysis

- Both bias and variance are very low for the 3 degree fit. The model is a better fit for the data.

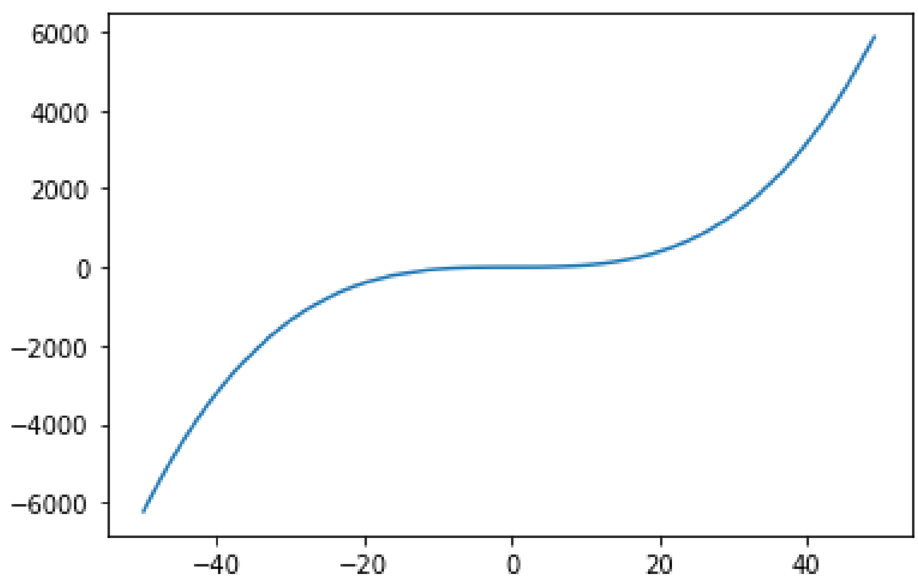
- For higher degrees (>3), both bias and variance increase. The training data has a lot of noise, hence they increase the bias and the variance both.

- Variance increases again for even higher degree polynomials (>6). The model is overfitting, and hence is not a good fit for the test data and is accounting for small deviations, making variance high.

3.4 Analysis of Data given

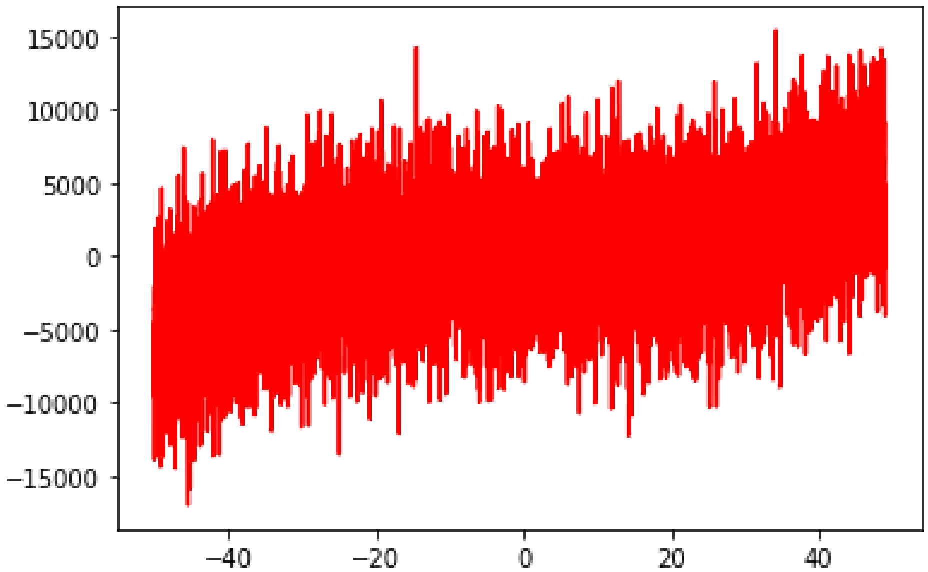
From the above observations, we can see that the test data is very noisy compared to the training data. Hence the bias and variance are both high at the linear fit, and then reach a minimum at the degree 3. The higher degree models try to acount for the noise and both bias and variance become higher.

3.4.1 Plot of Training Data



We see that a three degree fit is the best fit for the training data.

3.4.2 Plot of Test Data



We see that the test data is very noisy compared to the training data, which explains the high bias and variance.