Jiaming Li
CSI 5340 Intro to Deep Learning and Reinforcement Learning
Homework Exercise I
Sept 20, 2021

The homework is programmed by Python and mainly based on Numpy. The program analyzes three different factors, which are the size N of training dataset, the degree d of the model polynomial and noise variance σ^2 that may influence the result. The program consists of five parts: getData( ), which generates a dataset and is used to fit the regression model; getMSE( ), which computes the mean square error (MSE); fitData( ), which estimates the polynomial coefficients by fitting a given dataset to a degree-d polynomial; experiment( ), which takes as input the N, d and σ^2, and gets the output E_{in}, E_{out} and E_{bias}; and run the main code part.

The plot of getData( ) is shown below:

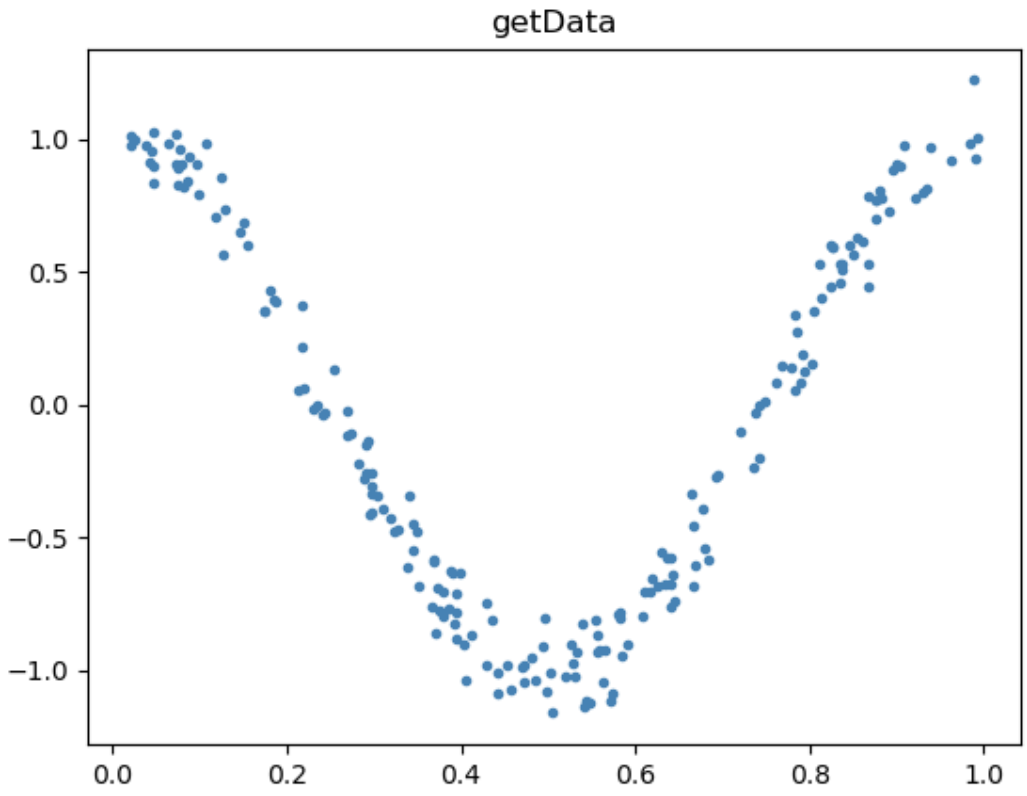

Figure 1: Scatter Plot of Getting Data

In addition to the three parameters mentioned above, the rest of parameters which has not given in the assignment are set as follows:

- Data test (data_test) = 2000

- Data bias (data_bias) = 2000
- Learning rate (learning_rate) = 0.01
- Epochs (epochs) = 1000
- Trials (M) = 50
- Lambda Value (lambda_value) = 1 (used for weight decay regularization)

Part I: Results without weight decay regularization

This part discusses three parameters which are Sample Size N, Model Complexity d and Noise Level Sigma and the results are based without weight decay regularization.

1. Result of Different Sample Size N

Except the parameters are given above, for Sample Size N, Model Complexity d and Noise Level Sigma are set as follows.

1.1 Parameters Settings

- Sample Size: $N \in \{2, 5, 10, 20, 50, 100, 200\}$
- Model Complexity: d = 10
- Noise Level (Variance): $\sigma^2 = 0.1$

1.2 Plot Show

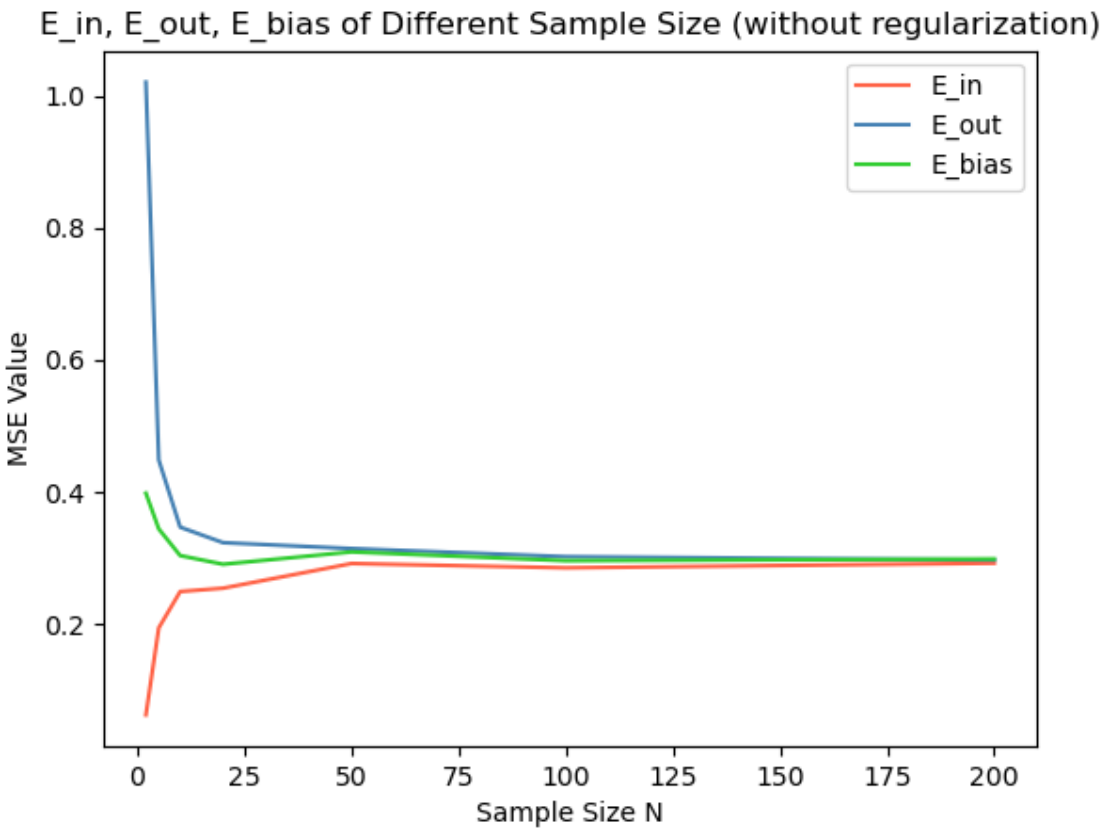

Figure 2: Result of Different Sample Size without Regularization

1.3 Conclusion

From figure 2 shows above, it is shown that both E_{out} and E_{bias} are decreasing rapidly until sample size N is greater than around 25. When sample size is at 50 or even above, E_{in} gradually flattens out, not rises, and approaches E_{out} and E_{bias}, all of which are in the same position. At this point, the conclusion can be drawn that the more dataset prepared, the better model will be generated since E_{out} and E_{in} has a fewer gap between them while the dataset is becoming more and more.

2. Result of Different Model Complexity d

2.1 Parameters Settings

- Sample Size: N = 100
- Model Complexity: $d \in \{0, 1, 2, 3, \dots, 19, 20\}$
- Noise Level (Variance): $\sigma^2 = 0.1$

2.2 Plot Show

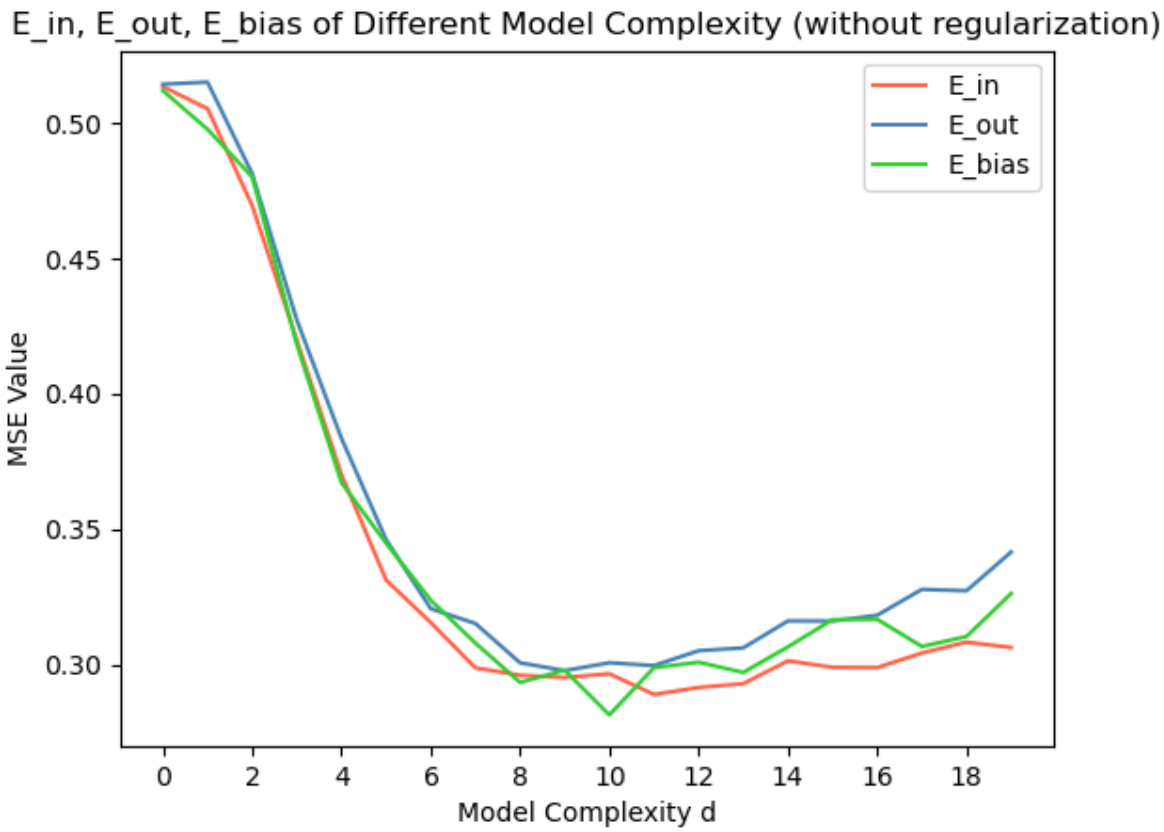

Figure 3: Result of Different Model Complexity without Regularization

2.3 Conclusion

What can be concluded by the plot shown above is that when complexity is at around 8 or

greater, the error (MSE) reaches the lowest peak. But after that, with the model complexity becomes higher, the error starts to change and does not stay stably.

3. Result of Different Noise Level Sigma

3.1 Parameters Settings

- Sample Size: N = 100
- Model Complexity: d = 10
- Noise Level (Variance): $\sigma^2 \in \{0.01, 0.1, 1\}$

3.2 Plot Show

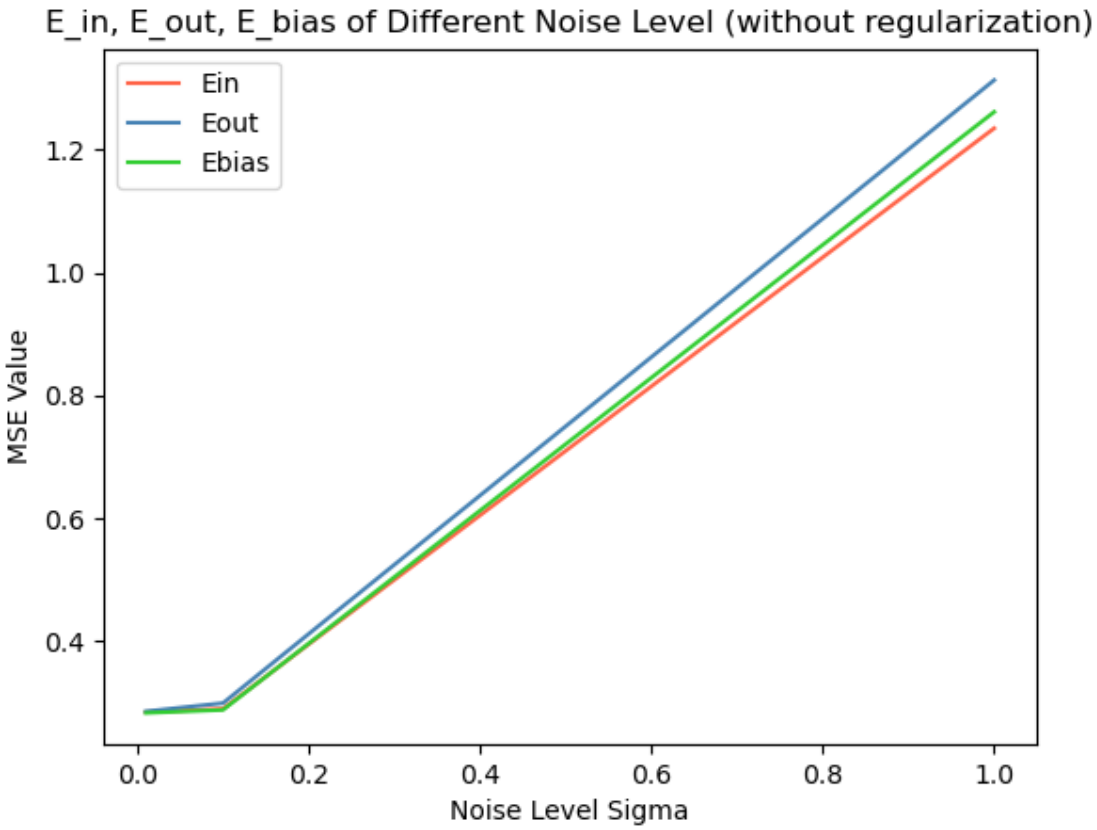

Figure 4: Result of Different Noise Level without Regularization

3.3 Conclusion

The conclusion is drawn straightforward from the picture since it is more likely a line. It can be seen roughly that sigma starts at 0.1, and as sigma grows, MSE value gradually rises. Moreover, the E_{out} line is always above the other two lines, and the value is always lager than E_{bias} and E_{in}.

Part II: Results with weight decay regularization

This part discusses three parameters which are Sample Size N, Model Complexity d and Noise Level Sigma and the results are based with weight decay regularization.

1. Result of Different Sample Size N

Except the parameters are given above, for Sample Size N, Model Complexity d and Noise Level Sigma are set as follows.

1.1 Parameters Settings

- Sample Size: $N \in \{2, 5, 10, 20, 50, 100, 200\}$
- Model Complexity: d = 10
- Noise Level (Variance): $\sigma^2 = 0.1$

1.2 Plot Show

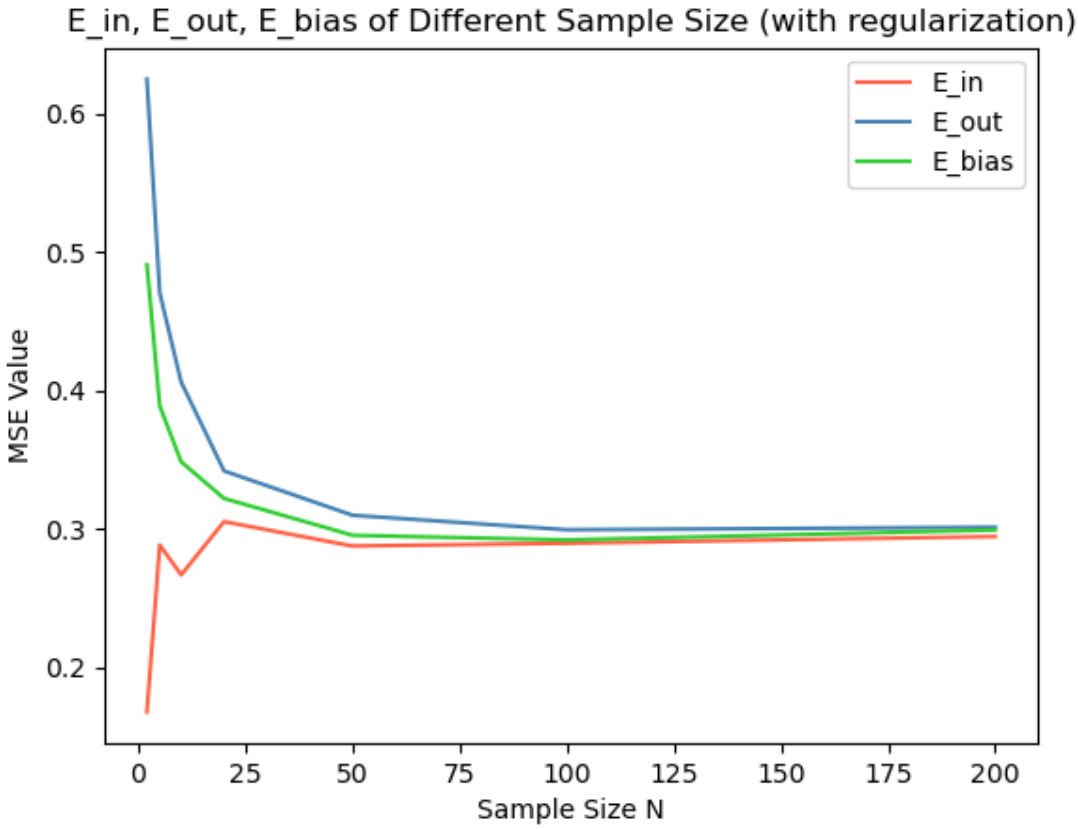

Figure 5: Result of Different Sample Size with Regularization

1.3 Conclusion

Figure 2 and Figure 5 look very similar since all three lines have the same trends. E_{out} and E_{bias} are decreasing, however, E_{in} is increasing before sample size reaches 25.

2. Result of Different Model Complexity d

2.1 Parameters Settings

- Sample Size: N = 100
- Model Complexity: $d \in \{0, 1, 2, 3, \ldots, 19, 20\}$
- Noise Level (Variance): $\sigma^2 = 0.1$

2.2 Plot Show

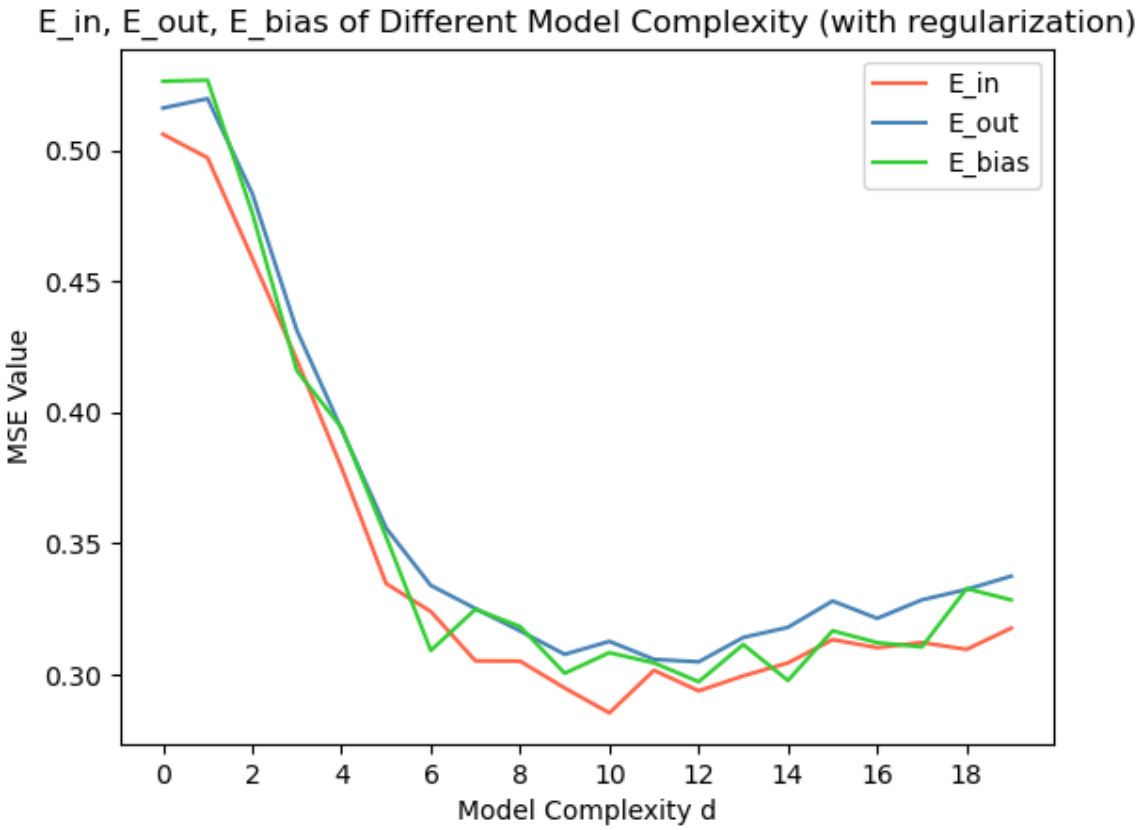

Figure 6: Result of Different Model Complexity with Regularization

2.3 Conclusion

Compare with Figure3, Figure 6 is similar to it. E_{out} line is always above the other two lines, and E_{in} has the smallest MSE value when three lines decrease rapidly. E_{bias} fluctuates between E_{in} and E_{out} after model complexity d is equal to 7.

3. Result of Different Noise Level Sigma

3.1 Parameters Settings

- Sample Size: N = 100
- Model Complexity: d = 10
- Noise Level (Variance): $\sigma^2 \in \{0.01, 0.1, 1\}$

3.2 Plot Show

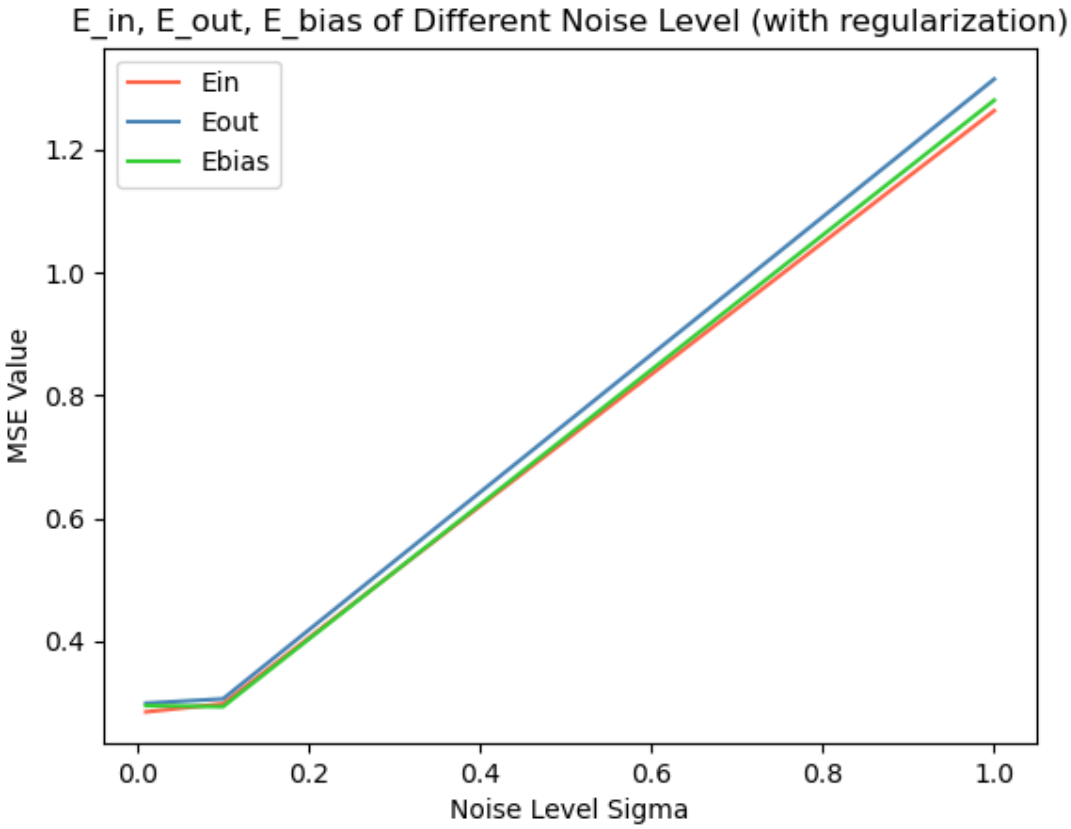

Figure 7: Result of Different Noise Level with Regularization

3.3 Conclusion

The same trend as the Figure 4 without regularization. As sigma grows, MSE value gradually rises. And the MSE value is the largest on E_{out} line while E_{in} is the smallest.

Part III: Results Comparison

This part compares the results with regularization and the results without regularization. Two plots with the same parameters setting but with/without regularization are showed together as follows.

1. Result of Different Sample Size N

Except the parameters are given above, for Sample Size N, Model Complexity d and Noise Level Sigma are set as follows.

1.1 Parameters Settings

- Sample Size: $N \in \{2, 5, 10, 20, 50, 100, 200\}$
- Model Complexity: d = 10
- Noise Level (Variance): $\sigma^2 = 0.1$

1.2 Plot Show

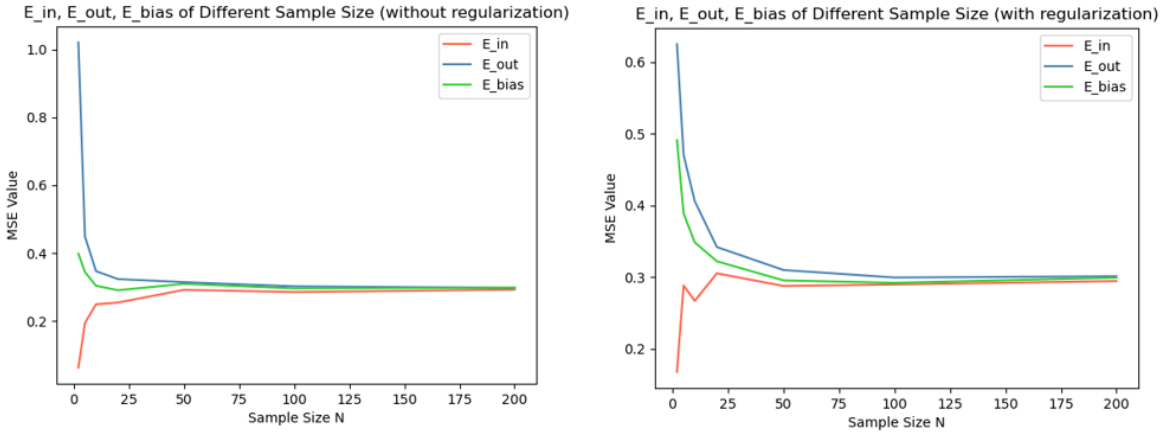

Figure 8: Comparison of the Result of Different Sample Size

1.3 Conclusion

Both E_{bias} and E_{out} of two plots decrease at the beginning and E_{in} increase, however, the MSE values of E_{bias} and E_{out} without regularization start with greater numbers than them with regularization. Finally, they all become closed to 0.3.

2. Result of Different Model Complexity d

2.1 Parameters Settings

- Sample Size: N = 100
- Model Complexity: $d \in \{0, 1, 2, 3, \ldots, 19, 20\}$
- Noise Level (Variance): $\sigma^2 = 0.1$

2.2 Plot Show

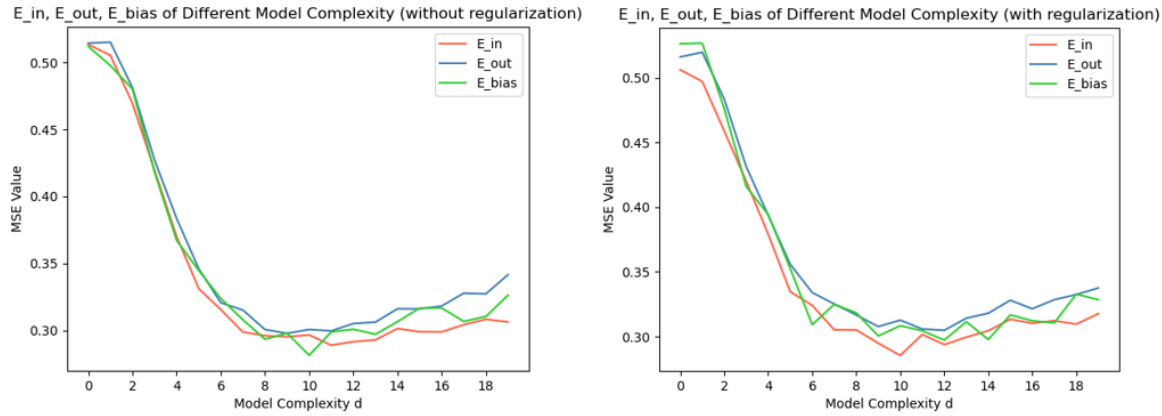

Figure 9: Comparison of the Result of Different Model Complexity

2.3 Conclusion

The two plots of Figure 9 show that the MSE value decreases when the model complexity d is

increased. E_{out} is the greatest while E_{in} has the smallest no matter which plot shows. And E_{bias} is at the middle between E_{in} and E_{out} if d is greater than 7.

3. Result of Different Noise Level Sigma

3.1 Parameters Settings

- Sample Size: N = 100
- Model Complexity: d = 10
- Noise Level (Variance): $\sigma^2 \in \{0.01, 0.1, 1\}$

3.2 Plot Show

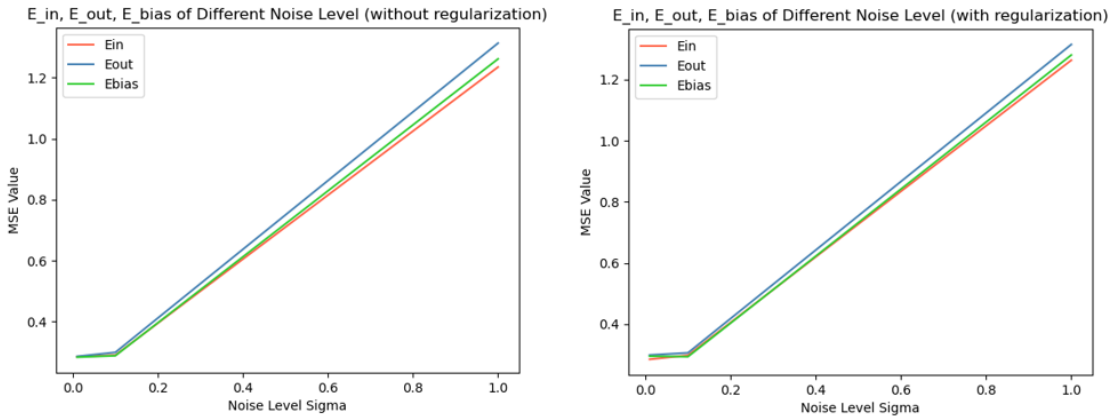

Figure 10: Comparison of the Result of Different Noise Level

3.3 Conclusion

As the noise level sigma increases, E_{bias}, E_{out} and E_{in} are rising as well. But the difference of their corresponding MSE values is smaller when with weight decay regularization, which means the result of the model would be more accurate and stable.