CSI 5138 Assignment 1

Yu Sun(8472921), Carleton University

September 2020

I mainly used Numpy to finish the gradient calculating and parameter updating. For the optimizer, I implemented the mini-batched SGD, which can transfer to GD with batch size equals to N, and to SGD with batch size equals to 1. After compared different models, I realised that best model should have minimal E_{gen}, which is balance between under-fitting and over-fitting. Moreover, using relative complexity model can solve the under-fitting problem but need to care about over-fitting issue. Besides, involving more data and using regularization can reduce the over-fitting problem. The regulation can improve the generalization of the model and decrease the complexity of the model.

1 Results without Regulation

In this section, I compared the models from three aspects: model complexity(d), sample size(N), noise level(σ), with the loss function of MSE .

1.1 Result of different model complexity

The result is plotted as Figure 1 shows, and the parameters I used list below:

Parameters Settings

- Complexity d: $d \in \{1, 2, ..., 20\}$;
- Dataset size N: $N = 100$;
- Variance σ: $\sigma = 0.1$;
- Number of trials M: 50;
- Learning rate Lr: 0.1;
- Batch size $batchsize$: 50;

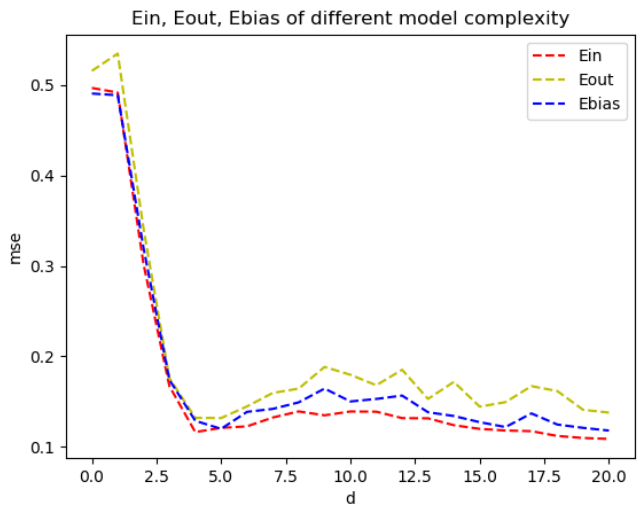


Figure 1: result of different model complexity

Conclusion

About E_{in}, it decreases with the increasing the model complexity, although sometimes it has the slight fluctuation. On the other hand, E_{out} decreases with the increasing the model complexity firstly, then it begins to increase, like the turning point of $d = 5$ in my experiment. Besides, E_{bias} has the similar trend with E_{out}.

The error gap ($E_{gen} = E_{out} - E_{in}$) reaches the minimal with $d = 5$, which indicates a threshold between under-fitting and over-fitting. Moreover, the model with $d = 5$ should be the best model in my case.

1.2 Result of different sample size

The result is plotted as Figure 2 shows, and the parameters I used list below:

Parameters Settings

- Complexity d: $d = 5$;
- Dataset size N: $N \in \{2, 5, 10, 20, 50, 100, 200\}$;
- Variance σ: $\sigma = 0.1$;
- Number of trials M: 50;
- Learning rate Lr: 0.1;
- Batch size $batchsize$: $== N$;

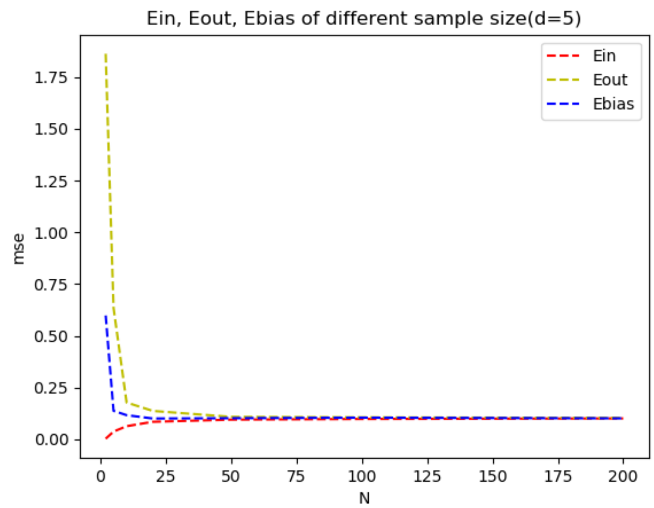
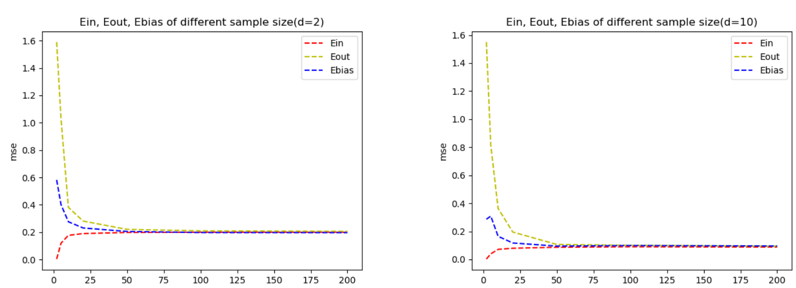


Figure 2: result of different sample size

Conclusion

With the data size increasing, E_{in} is increasing to a almost stable point. Meanwhile, E_{out}, E_{bias} are decreasing and converging to the same point. In another word, we can get smaller error gap between E_{out} and E_{in} with more data, which can learn a better model. The reason is that with small dataset, the model is usually over-fitting.

Besides, in this scenario, I also compare the different sample size of simple model and complex model. The convergence error point of complex model is smaller than that of simple model. So, with larger data, using complex model can get better result.



(a) simple model

(b) complex model

Figure 3: result of sample size with simple model and complex model

1.3 Result of different noise level

The result is plotted as Figure 4 shows, and the parameters I used list below:

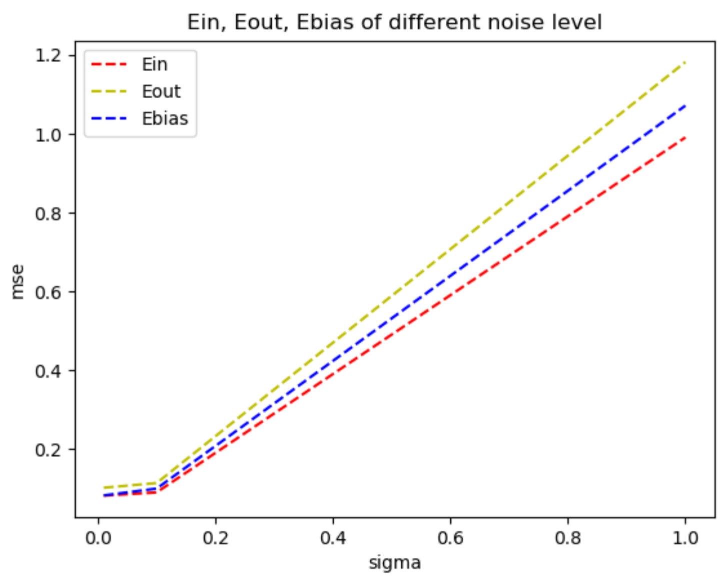


Figure 4: result of different noise level

Parameters Settings

- Complexity d: $d = 5$;
- Dataset size N: $N = 100$;
- Variance σ: $\sigma \in \{0.01, 0.1, 1\}$;
- Number of trials M: 50;
- Learning rate Lr: 0.1;
- Batch size $batchsize$: 50;

Conclusion

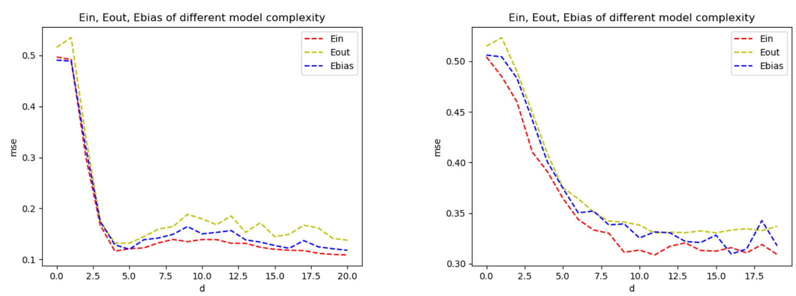
As we can see, with the increasing noise level, E_{in}, E_{out}, E_{bias} are increasing dramatically. Besides, the changes of E_{out}, E_{bias} are greater than E_{in}, which means the probability of over-fitting is larger with larger noise, and we need to do more to minimize the error gap.

2 Results with Regulation

In this section, I still compared from previous three aspects, and compared the result with or without regulation. Besides, the loss function add L2-Regularizer with $\lambda_{reg} = 0.1$, which will reduce the over-fitting issue. Except that, other parameters kept same as section 1.

2.1 Result of different model complexity

The result is plotted as Figure 5 shows, and the parameters I used list below:



(a) without regulation (b) with regulation

Figure 5: result of different model complexity with and without regulation

Parameters Settings

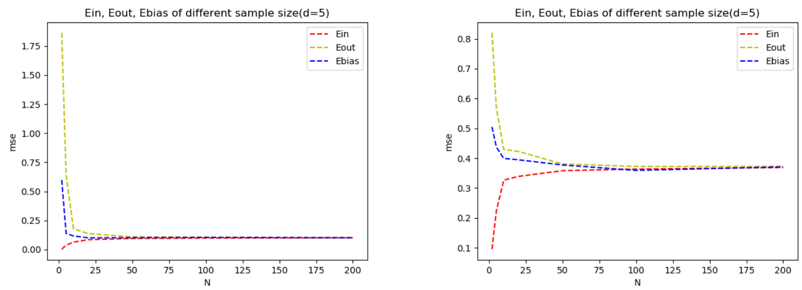
- Complexity d: $d \in \{1, 2, ..., 20\}$;
- Dataset size N: $N = 100$;
- Variance σ: $\sigma = 0.1$;
- Number of trials M: 50;
- Learning rate Lr: 0.1;
- Batch size $batchsize$: 50;
- $\lambda_{reg} = 0.1$;

Conclusion

Theoretically, regularization can reduce the gap between the E_{in} and E_{out}, in order to decrease the E_{out}. Through this way, it can benefit solving over-fitting problem and can also improve the generalization of the model. In my case, the error gap is slight smaller than that without regulation, but E_{out} is bigger than that without regulation. The reason is the model is not complex enough. If we use more complex model, we can get better result.

2.2 Result of different sample size

The result is plotted as Figure 6 shows, and the parameters I used list below:



(a) without regulation

(b) with regulation

Figure 6: result of different sample size with and without regulation

Parameters Settings

The result is plotted as Figure 6 shows, and the parameters I used list below:

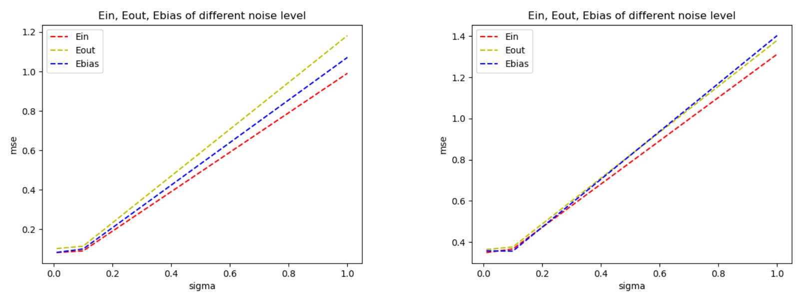
- Complexity d: $d = 5$;
- Dataset size N: $N \in \{2, 5, 10, 20, 50, 100, 200\}$;
- Variance σ: $\sigma = 0.1$;
- Number of trials M: 50;
- Learning rate Lr: 0.1;
- Batch size $batchsize$: $== N$;
- $\lambda_{reg} = 0.1$;

Conclusion

Those two scenarios have similar trend, which is that E_{out}, E_{bias} are drop and converge at the same point of E_{in}. But about the regulation, it need more time to get the best point.

2.3 Result of different noise level

The result is plotted as Figure 7 shows, and the parameters I used list below:



(a) without regulation

(b) with regulation

Figure 7: result of sample size with and without regulation

Parameters Settings

- Complexity d: $d = 5$;
- Dataset size N: $N = 100$;
- Variance σ: $\sigma \in \{0.01, 0.1, 1\}$;
- Number of trials M: 50;
- Learning rate Lr: 0.1;
- Batch size $batchsize$: 50;
- $\lambda_{reg} = 0.1$;

Conclusion

As we can see that, with the increasing noise level, E_{in}, E_{out}, and E_{bias} are increasing correspondingly. When noise level is bigger, the error gap of the regularization model is smaller, which means it can get more stable result and the model cannot be easy to over-fitting.