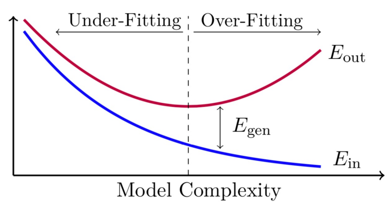
1 Results without regularization

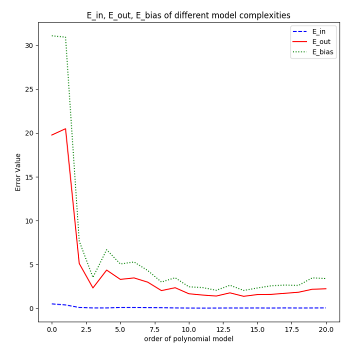
Since there are 3 changing parameters (N, d, and σ), we will focus on 1 parameter each time in the following subsections with fixing the other parameters. Other parameters are set as:
1. $\mathcal{D}_{test} = 2000$ and $\mathcal{D}_{bias} = 3000$;
2. $\alpha_{learning} = 0.1$ with decay 0.96 after each 100 steps;
3. $epoches = 1000$ for iteratively update loss function.

1.1 Result of different complexities

Complexity is researched at first, the result is plotted as Figure 1 shows,



(a) figure from textbook



(b) figure of test

Figure 1: Results of complexities.

1.1.1 Parameters Settings

- Complexity d: $d \in \{1, 2, ..., 20\}$;
- Dataset size N: $N = 50$;
- Variance σ: $\sigma = 0.1$.

1.1.2 Conclusion

As the experiment shows, the result of the experiment basically followed what we learnt from the lecture. The out-of-sample error starts extremely high when the model is too simple to cover the dataset. A significant drop occurs after the complexity of the polynomial reaches around 3 to 5. The error gap ($E_{gap} = E_{out} - E_{in}$)

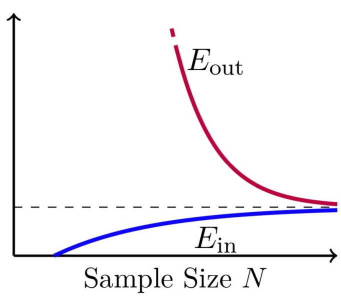
reaches the minimal when the order becomes around 12 15, and after that it goes up again. Thus, the over-fitting starts at order 12 15 when the dataset size is $N = 50$.

The in-sample error E_{in} shows a slightly decrease in the test figure. It is mainly because the variance is not that high and the number of training iteration is big, which makes the model learn well.

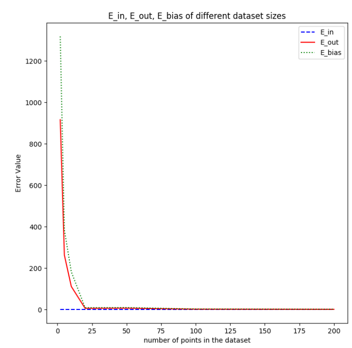
The bias error E_{bias} follows almost the same trend as the result of E_{out}.

1.2 Result of different sizes of dataset

Then, dataset size is researched, the result is plotted as Figure 2 shows,



(a) figure from textbook



(b) figure of test

Figure 2: Results of dataset size.

1.2.1 Parameters Settings

- Complexity d: $d = 5$;
- Dataset size N: $N \in \{2, 5, 10, 20, 50, 100, 200\}$;
- Variance σ: $\sigma = 0.1$.

1.2.2 Conclusion

As the experiment shows, the E_{out} and E_{bias} go all the way down and converge to a stable value just as the lecture slides indicate.

The E_{in} increase slightly step-by-step, and finally reach the same level with E_{out} and E_{bias}.

This implies that the larger the dataset, the smaller error gap we could get, and the better result the model could learn.

1.3 Result of different dataset variance

Then, dataset variance is researched, the result is plotted as Figure 3 shows,

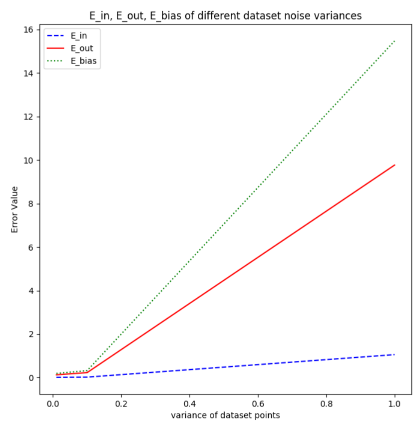


Figure 3: Results of dataset variance.

1.3.1 Parameters Settings

- Complexity d: $d = 10$;
- Dataset size N: $N = 200$;
- Variance σ: $\sigma \in \{0.01, 0.1, 1\}$.

1.3.2 Conclusion

As the experiment shows, all E_{in}, E_{out} and E_{bias} drametically increase during the whole changes with E_{bias} and E_{out} changing greater than E_{in}.

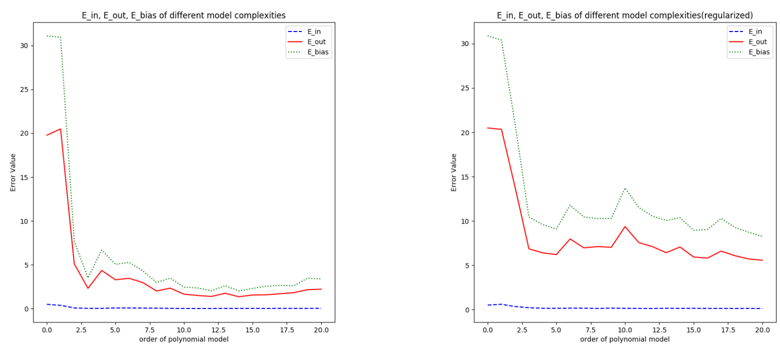
2 Results with regularization

Since we want to find the differences between regularized model and un-regularized model, in the following tests, only a new parameter is introduced with $\lambda_{reg} = 0.01$.

All other parameters are kept the same. So, we will skip the *paramters settings* part.

2.1 Result of different model complexities

The result comparison between complexity curve before regularization and after regularizaton is shown as Figure 4,



(a) Un-regularized (b) Regularized

Figure 4: Results of complexities.

2.1.1 Conclusion

As the experiment shows, all E_{in}, E_{out} and E_{bias} drametically increase during the whole changes with E_{bias} and E_{out} changing greater than E_{in}.