Experiment Descriptions

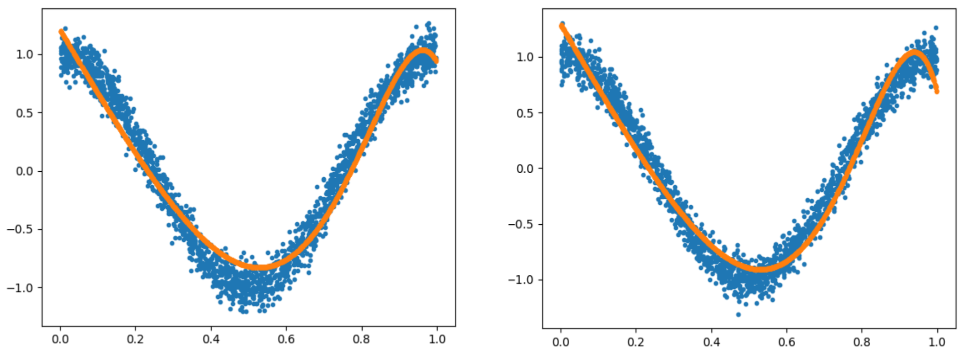
The purpose of this experiment is to examine the fitting and generalization of polynomials regression models via simulation. The experiment is carried out per the procedures described below.

1. Generate a dataset by the function getData(), which is the training data to train the model.
2. Iterate 2000 times to train the model according to the defined polynomials. This method is used to update the coefficients which are theta in the gradient descent formula.
3. Use different combinations of parameters (sample size N, model complexity d, and noise level σ) to get the results which are Ein, Eout, and Ebias. The results analysis will be based on these data.
4. Add the regularization (L2) to the loss function and observe the impact of weight decay.

Results Analysis

This section attempts to analyze different comparisons on the fitting and generalization of polynomials models, in relation to model complexity (d), sample size (N), and noise level (σ). Besides, comparisons are also studied as to whether the regularization has a positive correlation on models.

Firstly, it is essential to ensure that models designed will fit the dataset. In addition to model correctness, the learning rate and times of iteration also need to be determined. Besides, it is also important to ensure regularized models fit the data because the regularization term is in the loss function. For this part, the data generated by the cosine function with the gaussian random variable, and the predicted data by regression models are both plotted in the diagrams below. Two scenarios, one without the decay and one with the decay, are explored.



(left) Without the weight decay (right) With the weight decay

Figure1.

The orange line is the data generated from the given cosine function while the blue points are predicted by models. From these two figures it is discovered that both models can fit the data well.

This fitting process also helps determine the value of parameters to train models.. To be specific, the learning rate is 0.5 and iteration times are greater than 2000. From these results, it is concluded that the model can be used to perform subsequent experiments.

Determine the parameters

The figures below (Figure 2) show two experiments by the model. The difference between the two rows lies on whether the weight decay is included. The first-row experiments use the loss function without the regularization term. Each experiment first uses the largest sample size (N=200) and the smallest noise level (sigma=0.01) to observe the model complexity, which can be used to discover the rules of other parameters. The reason for choosing these two values is that a large sample size can train the model with enough data, with the assurance that the model will fit properly. After that, through using the degree of the model and the same sigma (sigma=0.01), we can study how the sample size will impact results. After these two parameters are determined, the results of different noise levels can be observed.

Comparing the model complexity, it can be noted that the small model complexity has a large error, which makes logical sense because the low model complexity means that the degree of polynomials is one or zero. The models are linear functions or vertical lines to the x-axis. The model does not fit the dataset and have a larger error than high-degree models. From the figure which is in the middle of first row, the error is small enough when the degree of polynomials is 6. The Ein and Eout of the degree after 6 approach to zero. The Ebias increase or decrease fairly slightly within a limited range. Therefore, d=6 is chosen as the value of the model complexity.

Comparing the sample size, the figure in the left of first row shows that the error will be smaller when the sample size becomes larger. In fact, when training models, a large dataset size usually means a more precise model. The Ein is small and the Eout is large at the low sample size. This is because the model we get may fit well in small sample sizes. However, this model cannot fit well in a new dataset which has magnitude difference of 1000 level with the training set.

Comparing the noise level, the figure in the right of first row shows that high noise level will cause the error becoming large, which is well expected. The high noise level means the training set can hardly be fitted. This eventually makes the model unable to fit the data. The Ein, Eout, and Ebias all increase with the noise level increase.

In the second row, the same experiments are conducted with the weight decay. It is noticed that the value of parameters determined without regularization can also work on the model with regularization. The errors are less significant between the figures in the top and bottom. Another figure will be used to observe the impact of weight decay.

From these figures, we can also find the Ebias errors are always larger than the Ein and Eout, which makes sense since the model is trained by a low dataset size. No matter how parameters change, the

Ebias remain in a constant distance with Ein and Eout.

To conclude, in this experiment it is determined that the appropriate values of parameters are Sample Size N=200, Model Complexity d=6, and Noise Level σ=0.001.

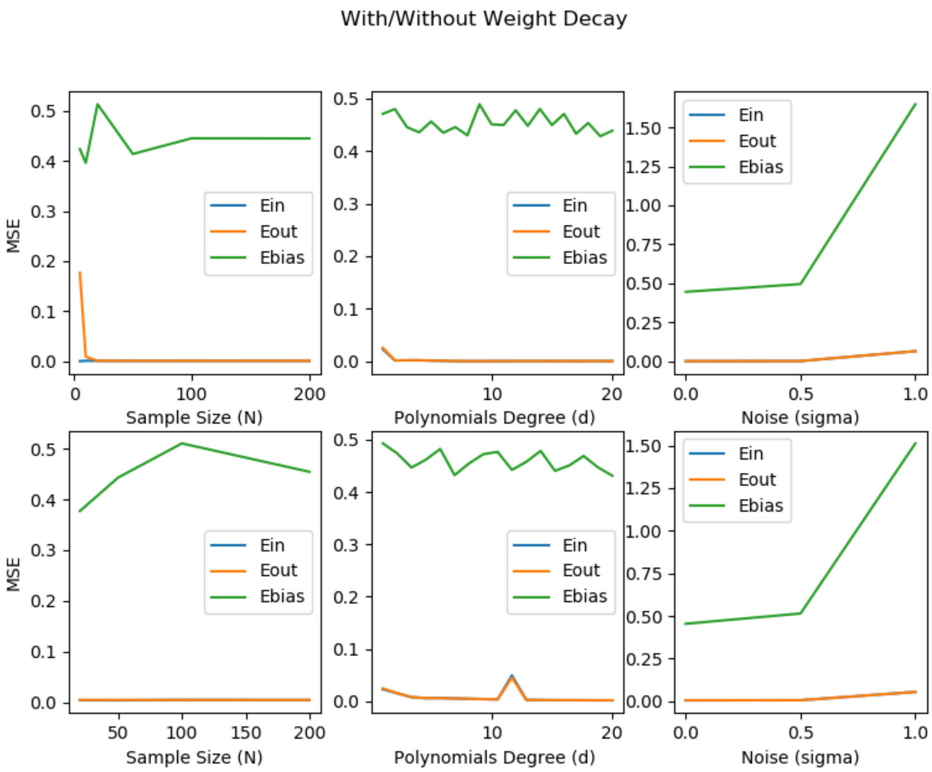


Figure 2

This last section will discuss the impact of weight decay. The parameters determined above are used. Ein, Eout, and Ebias will be compared respectively. The figures below show different Ein and Eout in same figure, and different Ebias are in same figures. The error calculated with weight decay is shown in dash line while the solid line represents the model without weight decay. The first column is the model trained in N=200, sigma=0.01, and the second column is the model of N=200 and d=6.

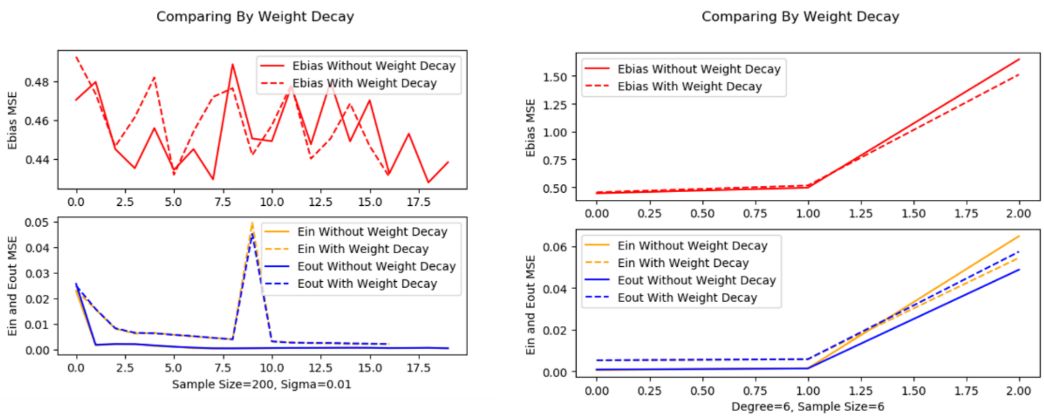


Figure 3

In addition, the figures indicate that the errors calculated with weight decay are always larger than those without weight decay, which is expected as the weight decay is a method used to reduce the over-fitting problem. It is common that over-fitting models usually have a small error. This means that weight decay shows a positive correlation with respect to models.

This summarizes the key findings through this experiment.