CSI5138 Assignment1

Name: Jiacheng Hou
Student Number: 300125708

In the assignment, I aim to fit a Polynomial Regression Model with training dataset (x, y) pairs generated from the equation $Y = \cos(2\pi X) + Z$, where X takes value in (0,1) and Z is a zero mean Gaussian random variable with variance σ^2.

I use Mini-batched SGD to train the model. I choose a batch size of 10 and a training epoch of 50. I calculate $\overline{E}_{in}$ using 50 trials of the training dataset and $\overline{E}_{out}$ using 50 trials of a separate large testing dataset, which contains 1000 data points. After that, I calculate the average obtained 50 polynomials over the 50 trials, and then generate another 1000 data points to fit the average polynomial model and calculate E_{bias}.

First of all, I plan to find the correlation between model complexity and $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias}. I find that when I choose the training dataset size from the list $[2, 5, 10, 20, 50, 100, 200]$, $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} all decrease with the polynomial degree increases from 0 to 20 degree. In Figura 1, I choose a training dataset size of 200, a noise variance of 0.0001, and polynomial degree $d \in \{0, 1, 2, ..., 20\}$.

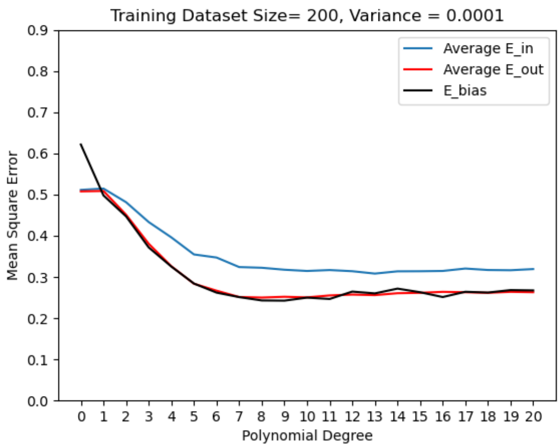


Figura 1: $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} changes with polynomial degree increases

Secondly, I plan to find the correlation between training dataset size and $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias}. I choose the training dataset size from the list $[2, 5, 10, 20, 50, 100, 200]$. I find that in the beginning, when the training dataset size is super small (< 10), the polynomial regression model fits the training dataset much better than the testing dataset. It means overfitting occurs when the training dataset size is too small. After that, with the increase of training dataset size, $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} all decrease. In Figura 2, I choose a polynomial degree of 8 and a noise variance of 0.0001, and training dataset $N \in \{2, 5, 10, 20, 50, 100, 200\}$.

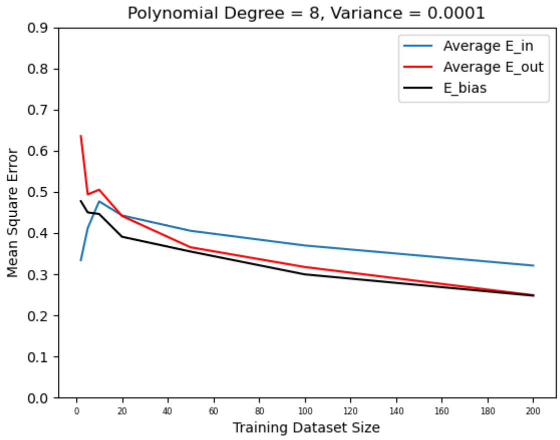


Figura 2: $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} changes with training dataset size increases

Thirdly, I plan to find the correlation between noise variance and $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias}. I find that with the increase of noise variance from the list $[0.0001, 0.01, 1]$, $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} all increase almost linearly. In Figura 3, I choose a training dataset size of 200 and a polynomial degree of 8, and noise variance $\sigma^2 \in \{0.0001, 0.01, 1\}$.

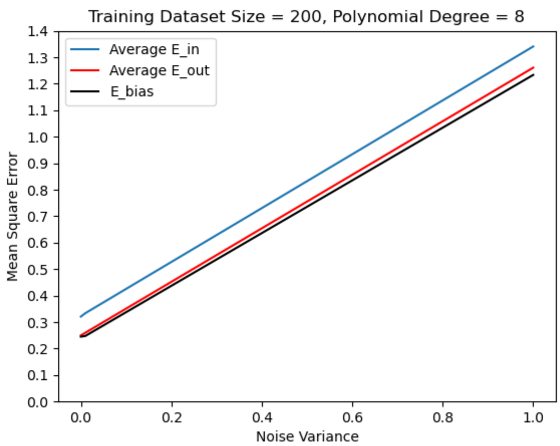
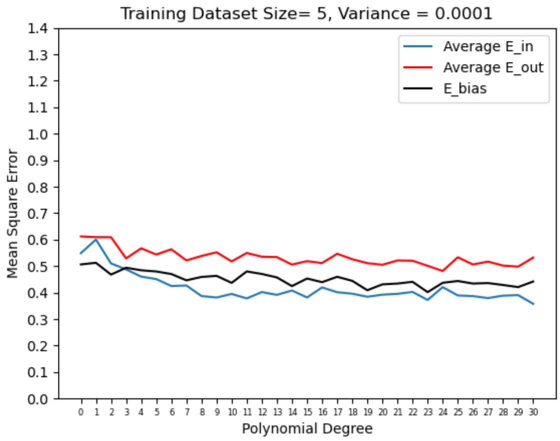


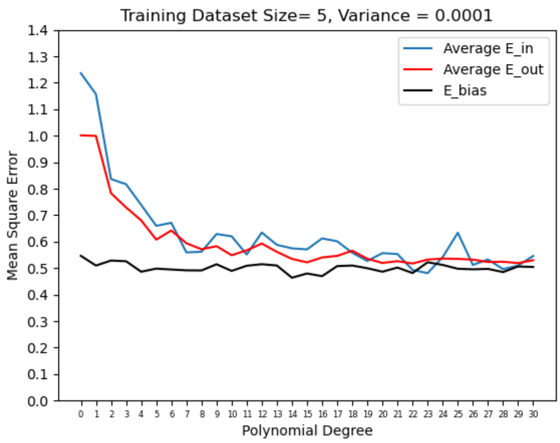
Figura 3: $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} changes with noise variance increases

Finally, I explored $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} values under the polynomial regression model without and with weight decay regularization, respectively. I find that when the training dataset size is super small (< 10), the model faces overfitting with the polynomial degree becoming larger. In Figura 4a, I choose a training dataset size of 5 and a noise variance of 0.0001, the difference between

$\overline{E}_{out}$ $\overline{E}_{in}$ is becoming larger and larger as the polynomial degree increases from 0 to 30. However, in Figura 4a, I add weight decay regularization to the previous model, the model doesn't overfit, and $\overline{E}_{in}$, $\overline{E}_{out}$ values always have little difference.



(a) Polynomial regression model without weight decay regularization



(b) Polynomial regression model with weight decay regularization

Figura 4: $\overline{E}_{in}$, $\overline{E}_{out}$ and E_{bias} under polynomial regression model