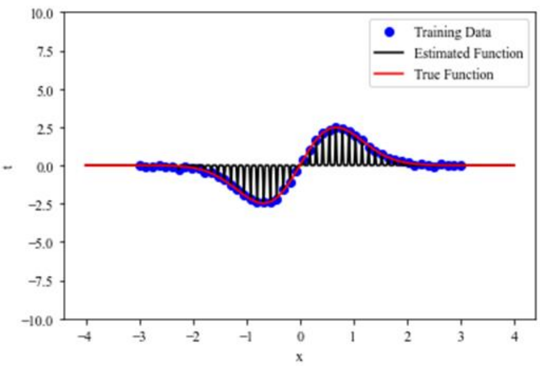
Assignment 02

Solution by Kaifa Lu

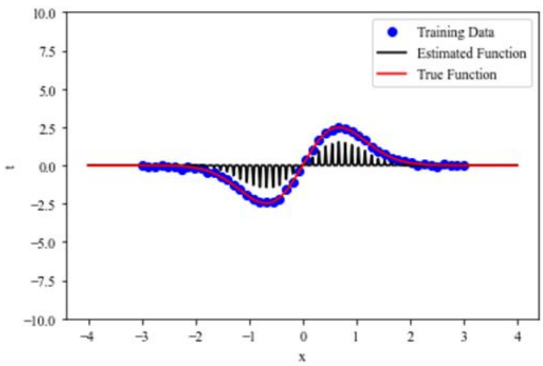
October 11, 2021

Investigation

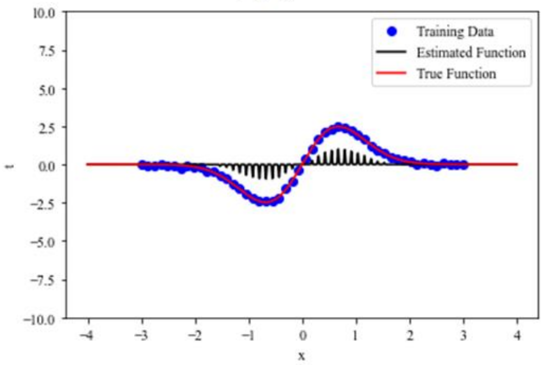
(1) Plots when M=50, s=0.01, $\lambda_1=\lambda_2=0.5$, learning rate=0.001, and α is varying from 0 to 10 with a step size of 1



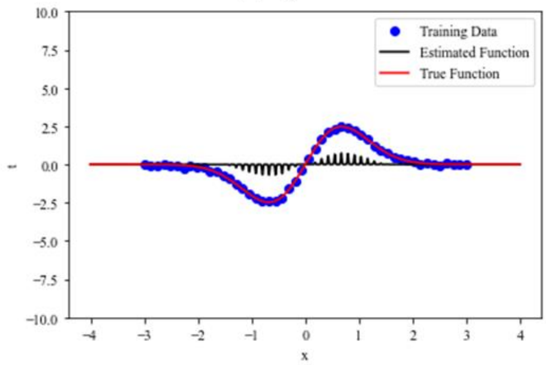
$\alpha=0$, $\lambda_1=\lambda_2=0.5$



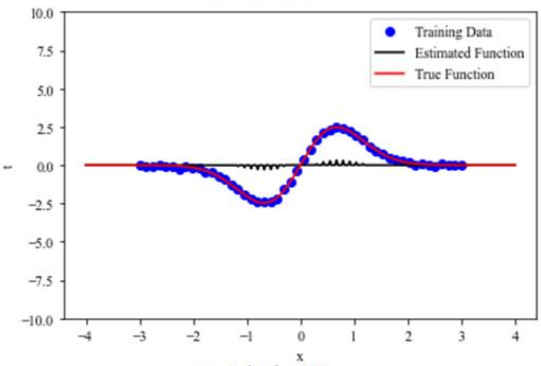
$\alpha=1$, $\lambda_1=\lambda_2=0.5$



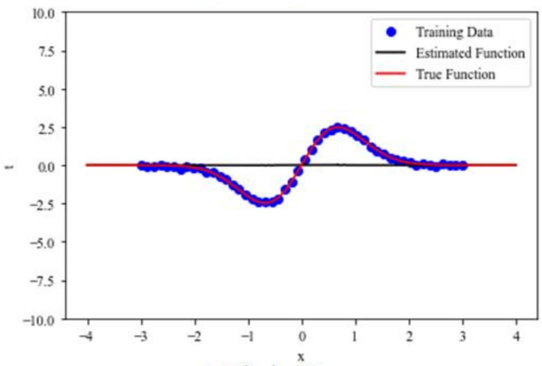
$\alpha=2$, $\lambda_1=\lambda_2=0.5$



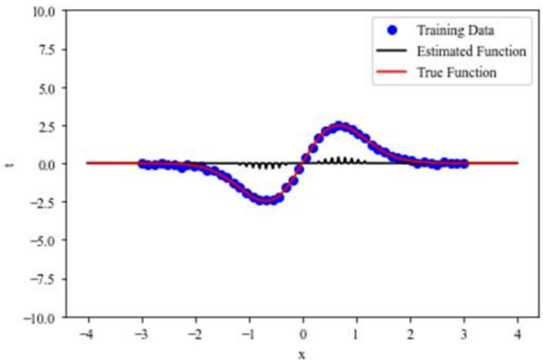
$\alpha=3$, $\lambda_1=\lambda_2=0.5$



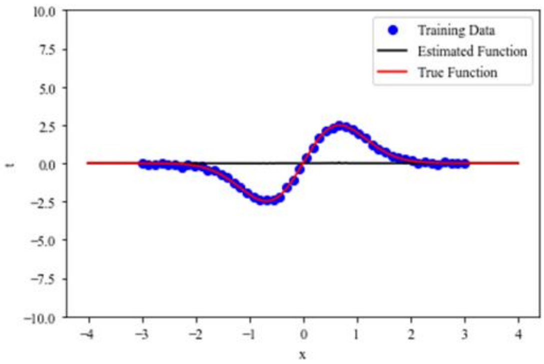
$\alpha=4$, $\lambda_1=\lambda_2=0.5$



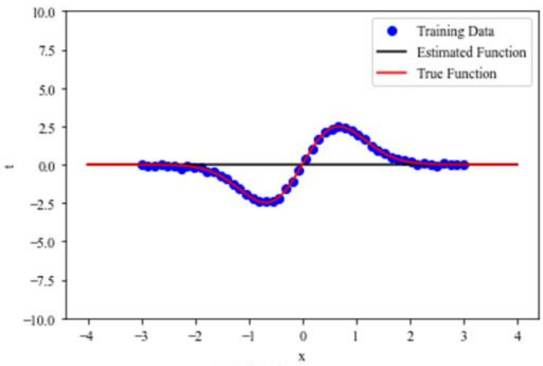
$\alpha=5$, $\lambda_1=\lambda_2=0.5$



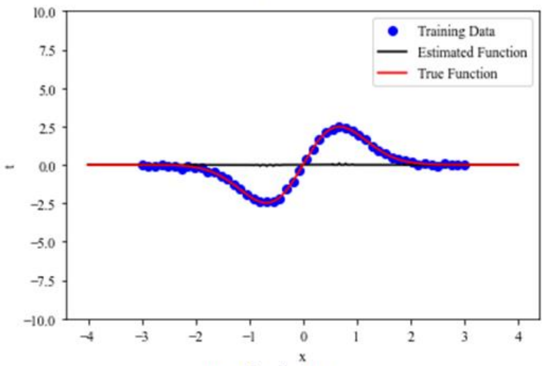
α=6, λ_1=λ_2=0.5



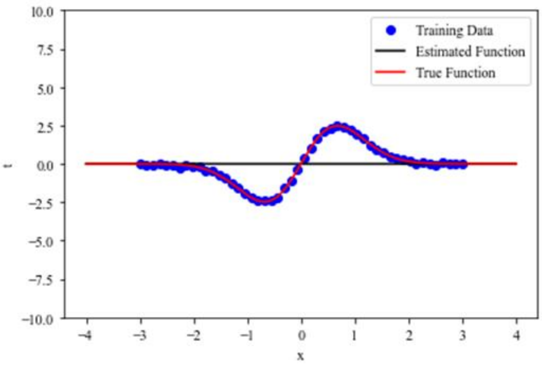
α=7, λ_1=λ_2=0.5



α=8, λ_1=λ_2=0.5



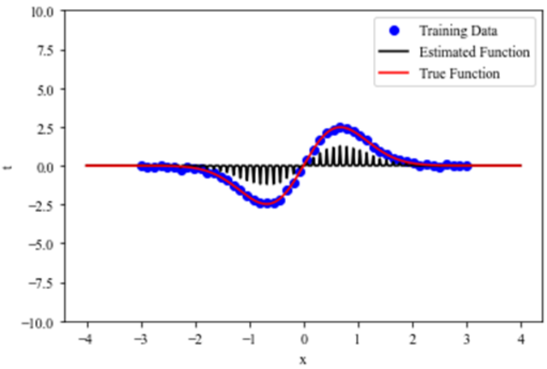
α=9, λ_1=λ_2=0.5



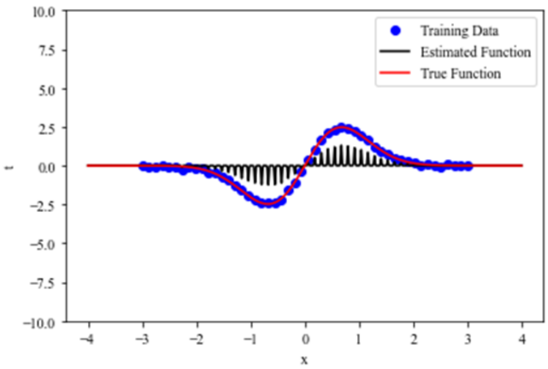
α=10, λ_1=λ_2=0.5

Figure 1: Prediction when λ_1=λ_2=0.5 and α varies from 0 to 10

(2) Plots when M=50, s=0.01, α=1, learning rate=0.001, and λ_1 is varying from 0 to 1 with a step size of 0.1



α=1, λ_1=0



α=1, λ_1=0.1

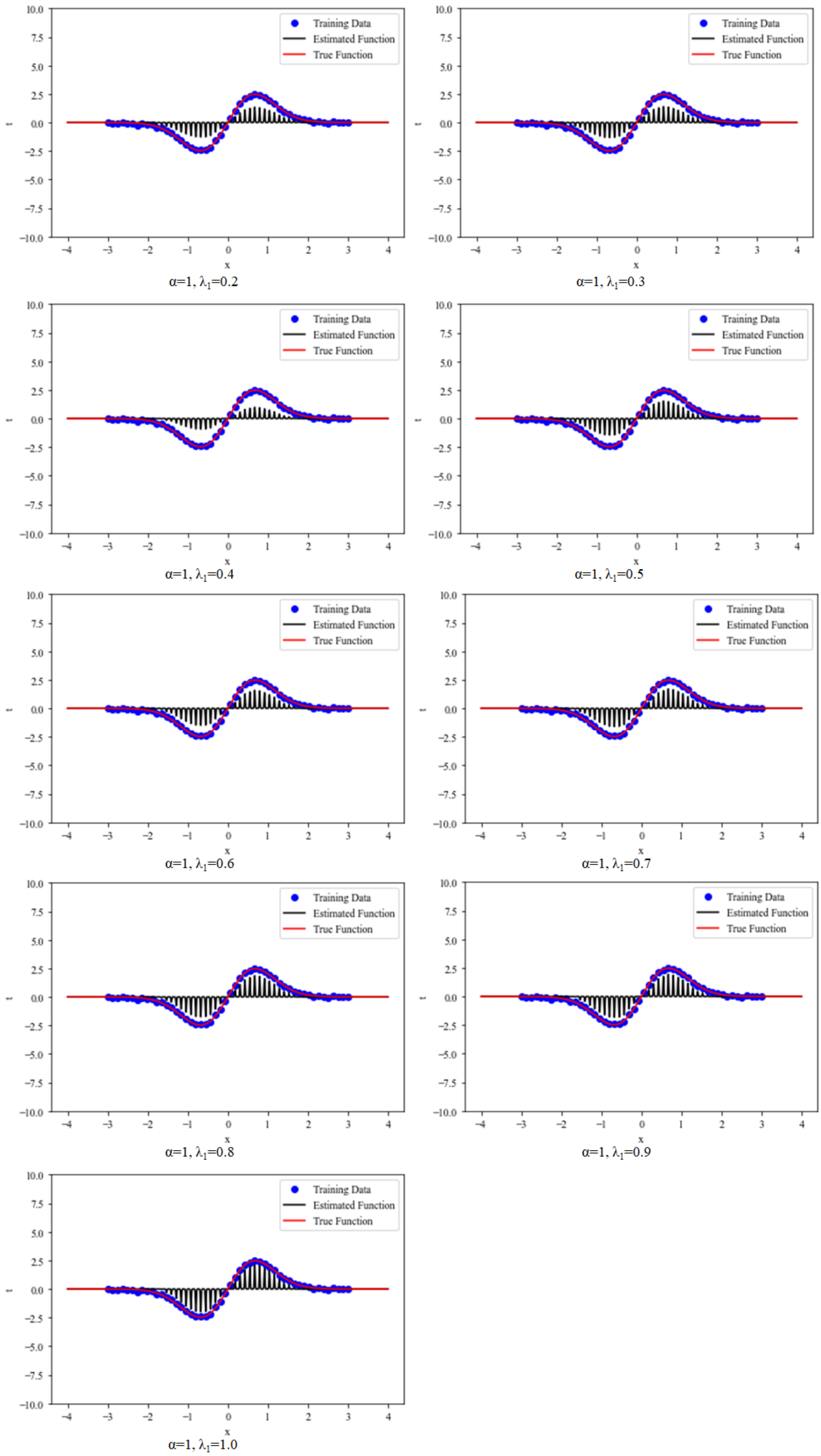
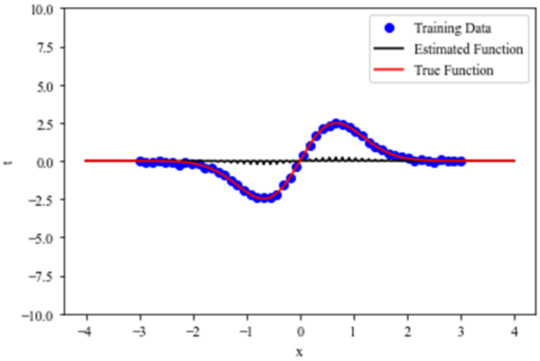
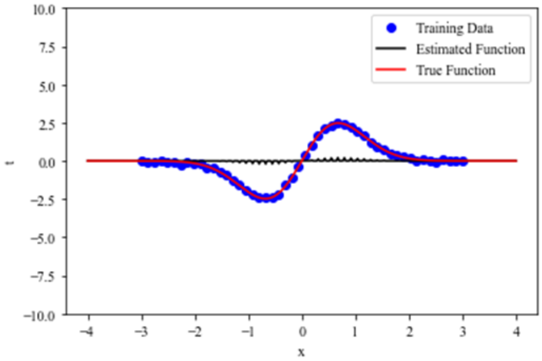


Figure 2: Prediction when α=1 and λ_1 varies from 0 to 1

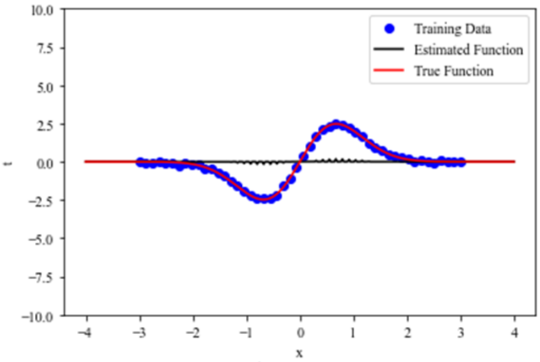
(3) Plots when M=50, s=0.01, α=5, learning rate=0.001, and λ_1 is varying from 0 to 1 with a step size of 0.1



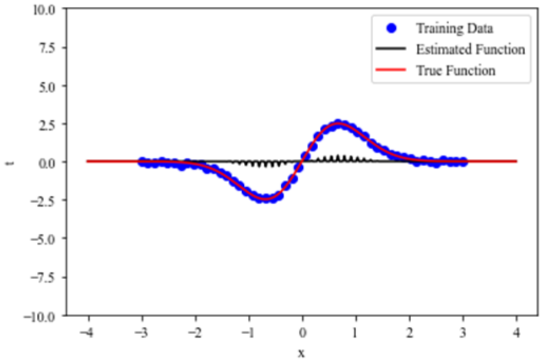
α=5, λ_1=0.0



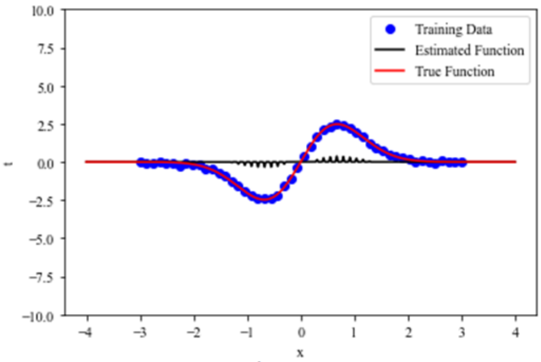
α=5, λ_1=0.1



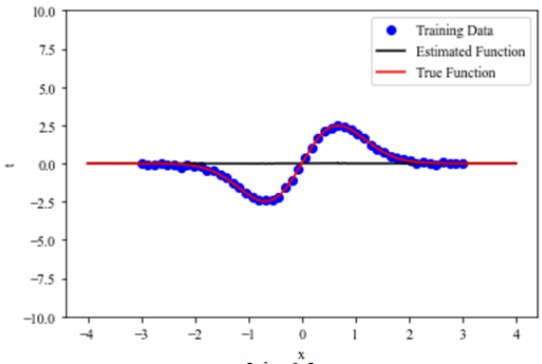
α=5, λ_1=0.2



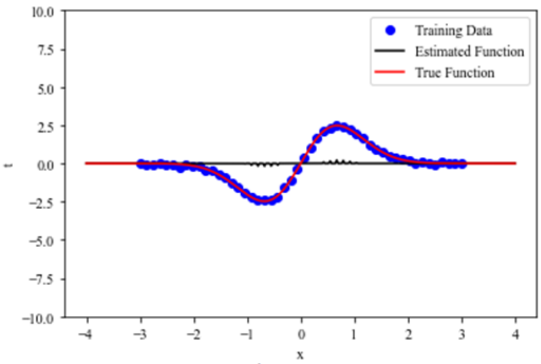
α=5, λ_1=0.3



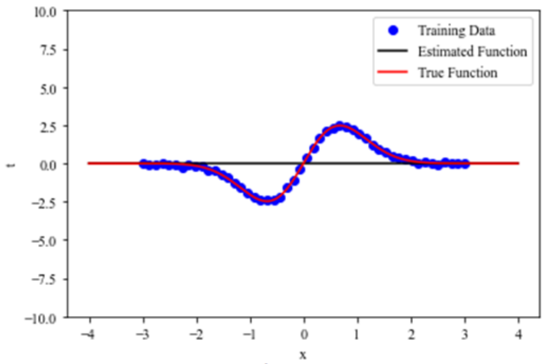
α=5, λ_1=0.4



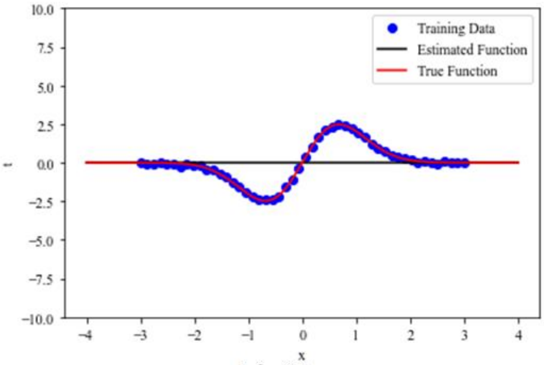
α=5, λ_1=0.5



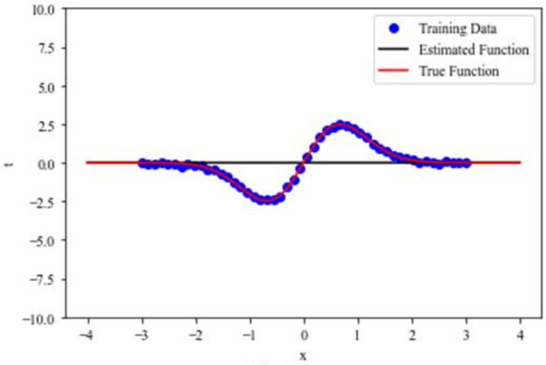
α=5, λ_1=0.6



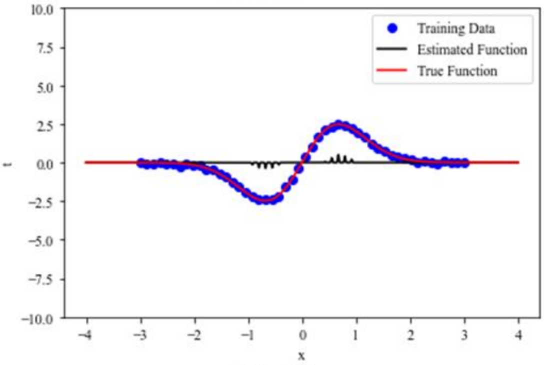
α=5, λ_1=0.7



α=5, λ_1=0.8



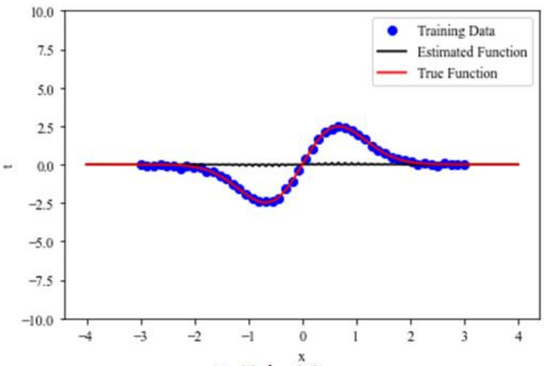
α=5, λ_1=0.9



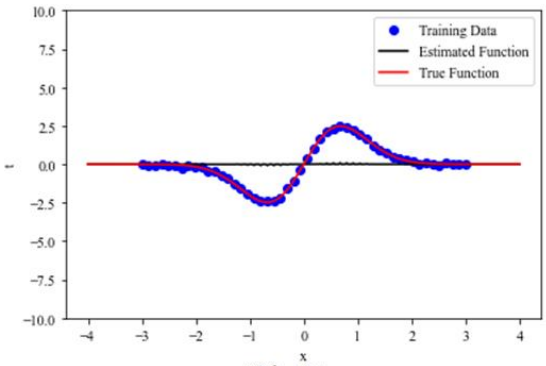
α=5, λ_1=1.0

Figure 3: Prediction when α=5 and λ_1 varies from 0 to 1

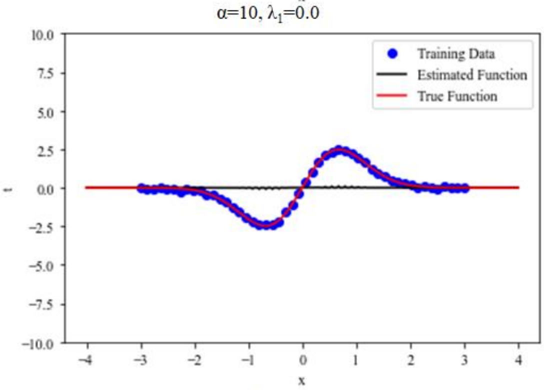
(4) Plots when M=50, s=0.01, α=10, learning rate=0.001, and λ_1 is varying from 0 to 1 with a step size of 0.1



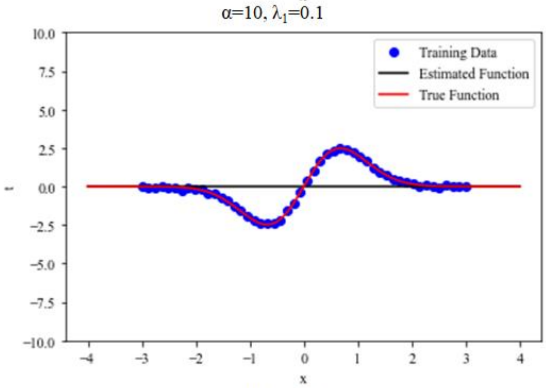
α=10, λ_1=0.0



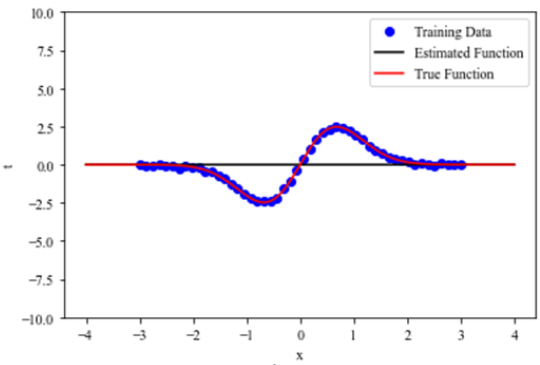
α=10, λ_1=0.1



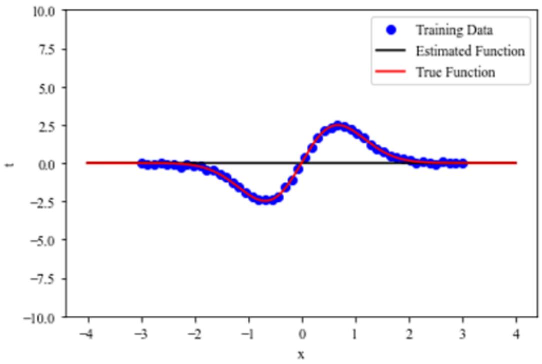
α=10, λ_1=0.2



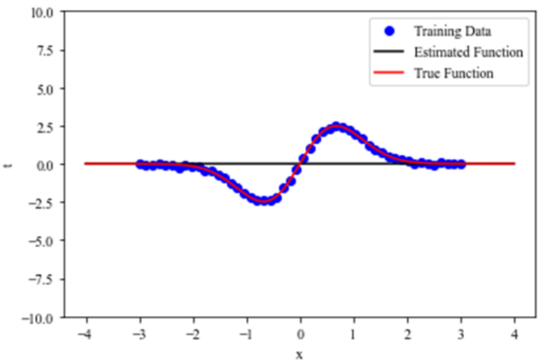
α=10, λ_1=0.3



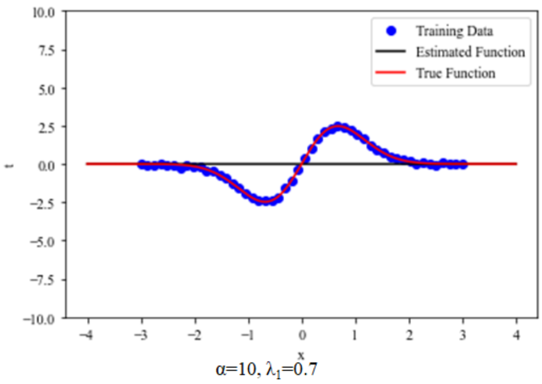
α=10, λ_1=0.4



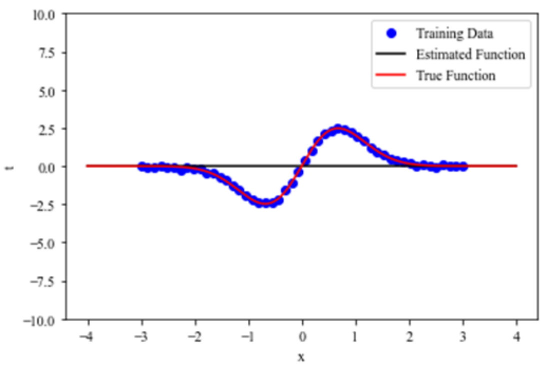
α=10, λ_1=0.5



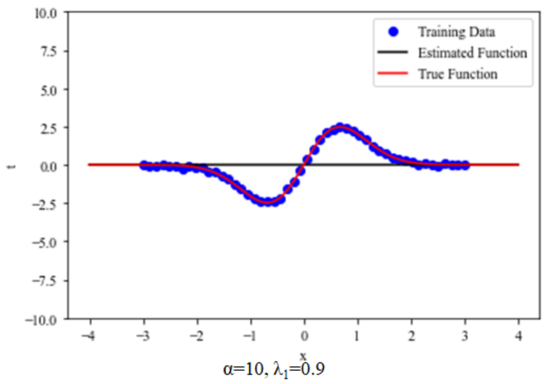
α=10, λ_1=0.6



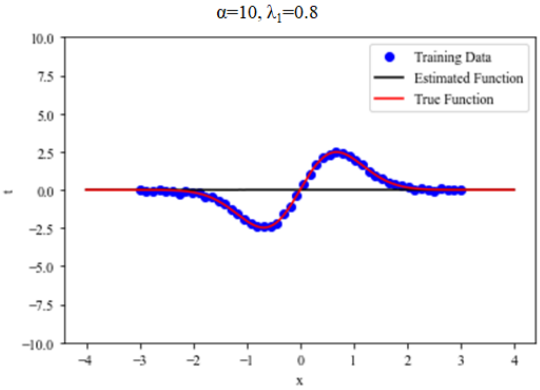
α=10, λ_1=0.7



α=10, λ_1=0.8



α=10, λ_1=0.9



α=10, λ_1=1.0

Figure 4: Prediction when α=10 and λ_1 varies from 0 to 1

(5) 3D mesh showing the absolute error on the training data when α is varying from 0 to 10 with a step size of 1 and λ_1 is varying from 0 to 1 with a step size of 0.01

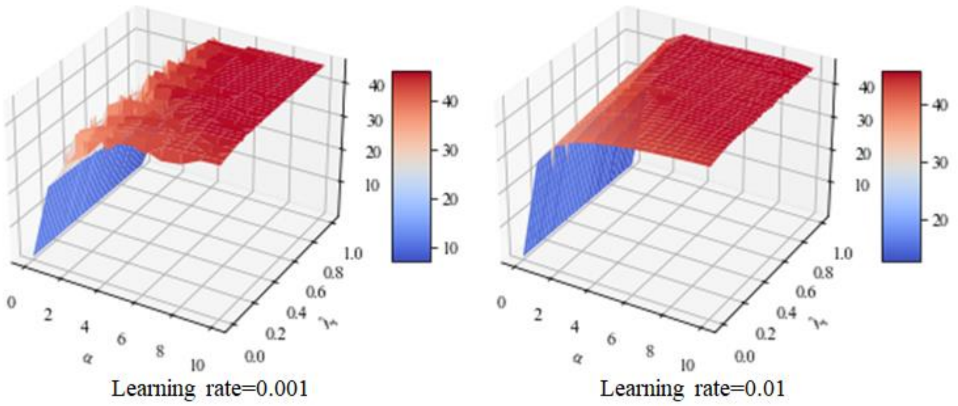


Figure 5: The absolute error behavior on the training data when α and λ_1 vary

(6) 3D mesh showing the absolute error on the validation data when α is varying from 0 to 10 with a step size of 1 and λ_1 is varying from 0 to 1 with a step size of 0.01

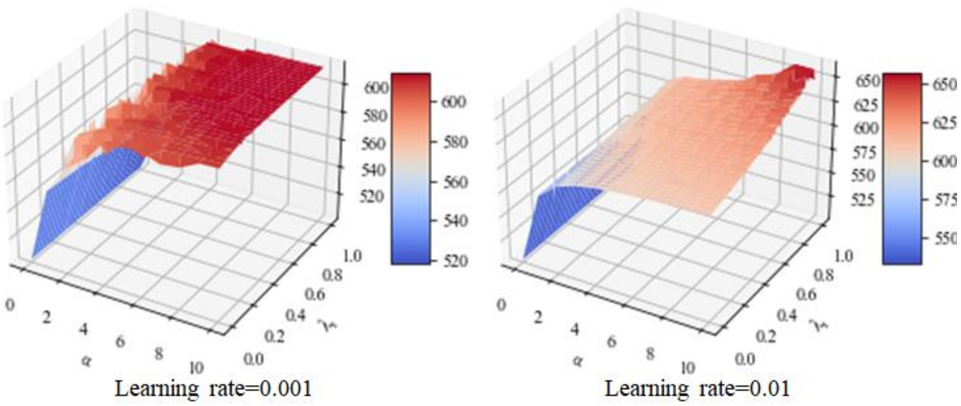


Figure 6: The absolute error behavior on the validation data when α and λ_1 vary

(7) 3D mesh showing the absolute error on the testing data when α is varying from 0 to 10 with a step size of 1 and λ_1 is varying from 0 to 1 with a step size of 0.01

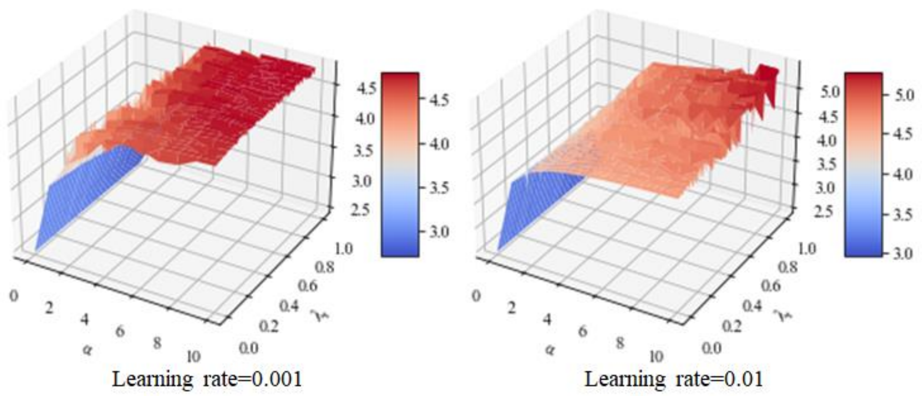


Figure 7: The absolute error behavior on the testing data when α and λ_1 vary

Discussion

Question Responses

1) **Based on your visualizations of absolute error above, what is the best combination of hyperparameter values using the full training data set. What is the best combination of hyperparameter values using the validation set? What is the best combination of hyperparameter values using the test set? Why or why not are the α, λ_1, and λ_2 values identified in here the same across all three data sets?**

A: As shown in Figure 5-7, the best combination of hyperparameters α and λ_1 across the training, validation and testing data sets is the same (α=0) when α is varying from 0 to 10 and λ_1 is varying from 0 to 1. That is to say, when α=0, the three data sets reach the least absolute error between the predicted and real values, compared with other combinations of α and λ_1. This indicates that multiple linear regression using radial basis functions without any regularization item show better prediction performances on the three data set than those with an Elastic Net regularization. This is mainly due to the fact that there is no overfitting phenomenon found on the training data set when performing the multiple linear regression using radial basis functions. Figure 8 shows the prediction on the three data sets when α=0.

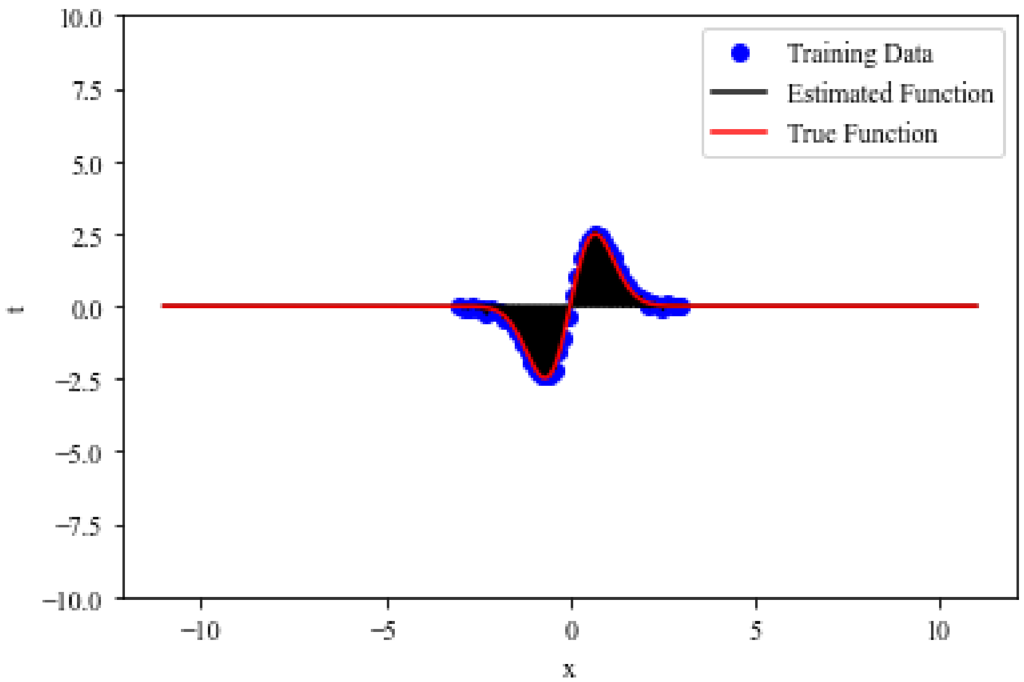


Figure 8: Prediction on the three data set when α=0

For multiple linear regression using RBFs with an Elastic Net regularization (α≠0), the best combination of hyperparameters α and λ_1 is α=1 and λ_1=0.96 using the training data set, α=1 and λ_1=1 using the validation data set, α=1 and λ_1=1 using the testing data set, respectively. Overall, the absolute errors on the three data sets show a gradually rising trend as α increases from 0 to 10, and the hyperparameter α exerts a greater influence on the absolute errors of the three data sets than the hyperparameter λ_1. Furthermore, multiple linear regression using RBFs with more L2 regularization components generally exhibits better prediction performances than that with more L1 regularization components.

2) **In this assignment, what are all the controlled and uncontrolled parameters that need to be considered when designing an experiment to evaluate the model? Suppose you would like to determine the expected performance of this model on unseen test data and reliability of that performance. Given these parameter settings, how would you design an experiment to determine these expected performance metrics? Carry out this experiment and report what you found to be the expected performance and reliability of that performance. (Part of this question is defining how you might measure and report "expected performance" and "reliability of performance")**

A: When designing an experiment to evaluate the model, the controlled parameters that need to be considered include three hyperparameters α, λ_1, λ_2, and M (the number of features/RBFs), s (the standard deviation of RBFs) and learning rate, while the uncontrolled parameter that need to be considered is the selection of centers of the features/RBFs.

The expected performance of the model on unseen testing set can be determined by the following evaluation indices: Mean Absolute Error (MAE), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE), etc. MAE and MSE denote the deviation between the real and predicted values of the model, however, MAE and MSE are easily enlarged by some outliers that can show significant deviations. Hence, MAPE is introduced to describe the model performance, which is one more robust evaluation index than MAE and MSE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - t_i|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - t_i)^2$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - t_i}{t_i}\right|$$

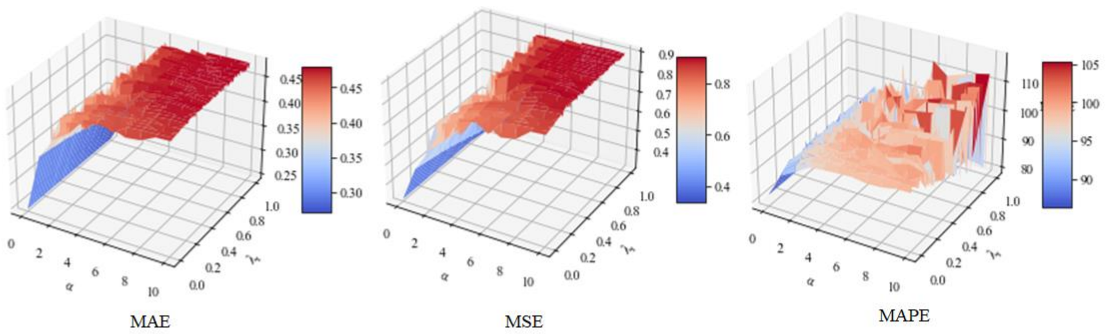


Figure 9: The error behaviors on the unseen testing data when α and λ_1 vary

To test the reliability of the model performance, several experiments were conducted by varying the learning rate of weight from 0.001 to 0.01 to explore how the above evaluation indices (MAE, MSE and MAPE) change. As shown in Figure 10, when the learning rate of weights varies from 0.001 to 0.01, the MAE, MSE and MAPE of the model on unseen testing data sets almost present no significant variations. Furthermore, as the α and λ_1 vary, the three evaluation indices when learning rate= 0.001 show the nearly same changing trends as those when learning rate=0.01. Only when α is approaching 10, the learning rate exerts a minor influence on the changing trends of the MAE, MSE and MAPE on unseen testing data sets. This suggests that the model basically demonstrates better reliability in the prediction performance. It is also worth mentioning that when learning rate=0.01, the model takes more training time (nearly double) to acquire the optimal weights than that when learning rate=0.001, particularly for a larger α value.

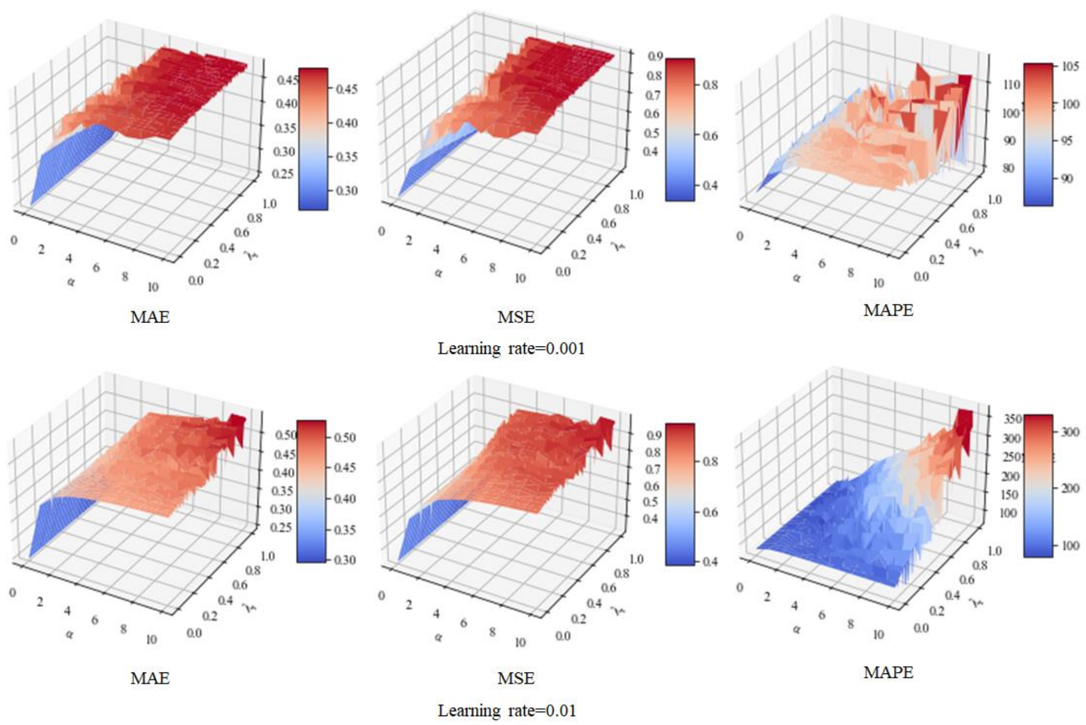


Figure 10: The error behaviors of learning rate=0.001 and 0.01 on the unseen testing data when α and λ_1 vary