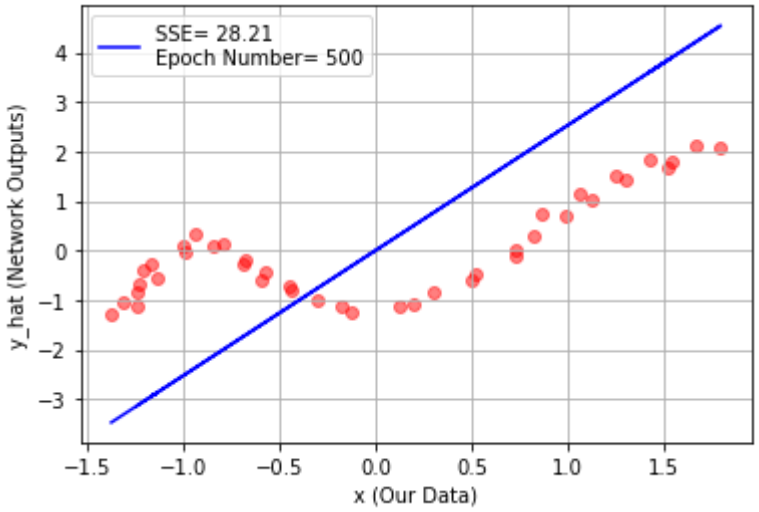
GE461 - Introduction to Data Science
Project 3 Report
Mustafa Yaşar
21702808

Question a.

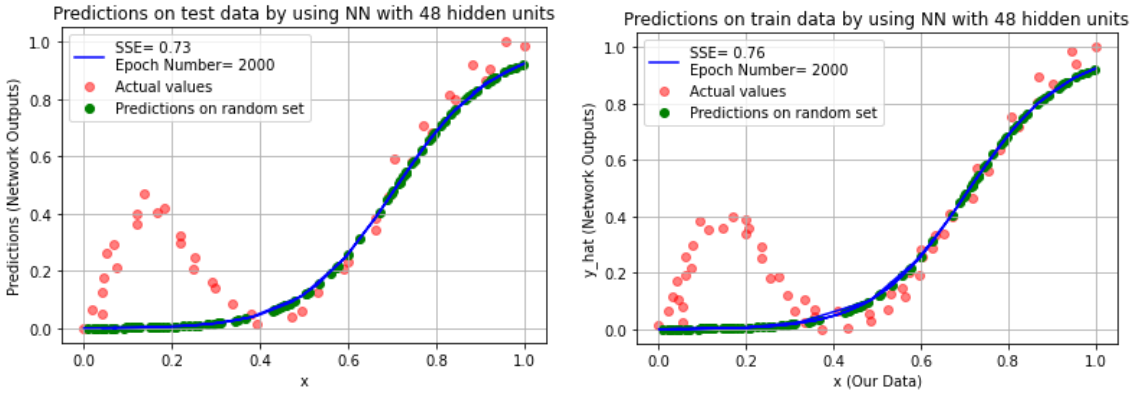
- It is not sufficient to use a linear regressor because the training data is curved and there is no single line that can explain most of the variance in the dataset. Therefore, it is required to use an artificial neural network with at least 1 hidden layer so that the model will be able to fit to curves.
- In order to determine hyper parameters such as learning rates and number of epochs, different types of these hyper parameters have been checked manually and the hyper parameters which give the least error have been chosen to train the ANN.
- When the error does not change, the change becomes statistically insignificant, or the error increases, the model can stop learning to decrease the computational burden and increase efficiency. For instance, while training a linear regressor without a hidden layer, the error when the epoch is 357 is 28.19. However, when the epoch is 500, the error becomes 28.33. Therefore, determining a good epoch number is important for the model.
- Normalization is one of the most significant aspects of the model training process.
    - While training a linear regressor without a hidden layer, the dataset has been standardized so that the mean becomes 0 and the standard deviation becomes 1.
    - While training a linear regressor with a hidden layer, min max normalization has been used because of the fact that the sigmoid function scales every input to 0-1 range. If we have used the

The obtained plot when the NN without hidden layer was used is as follows:



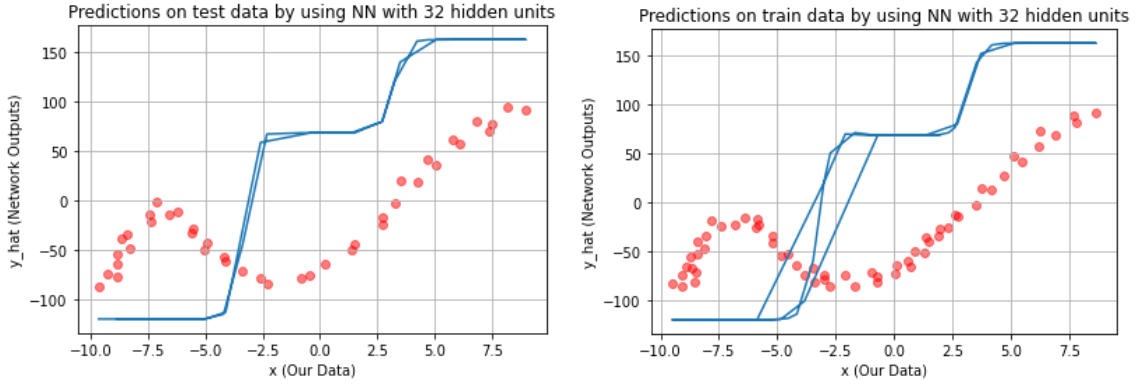
Question b.

After normalizing the dataset (both training and test sets), the best possible model contains 48 hidden units, the learning rate is 0.0001, and epoch count is 2000. The error obtained is 0.76. When the best hyper-parameters was used on the test and the train data, the following plots were obtained. The blue lines represent the predictions made by the neural network.



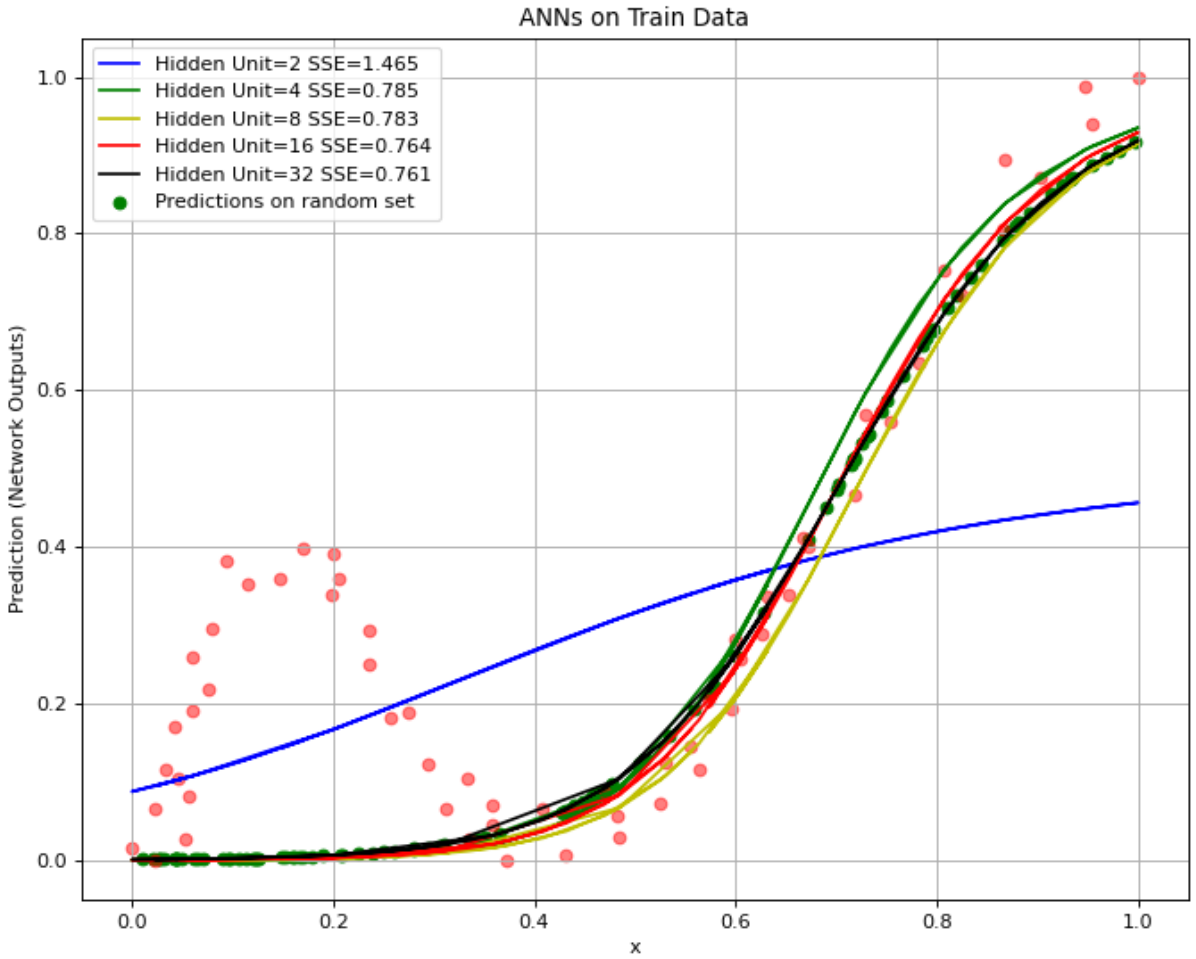
ANN Used: ANN with 1 hidden layer with 48 hidden units
Learning rate: 0.0001
Range of initial weights: (0-1)
Number of epochs: 2000
Is normalization used: YES
Training Loss: 0.75776
Test Loss: 0.73364

When the dataset was **not** normalized, the following plots were obtained.



Question c.

- First of all, the best learning rate for each ANN has been determined. When the number of hidden units is 2, 4, 8, 16, and 32, the best learning rate for each ann is 0.5, 0.5, 0.5, 0.5, and 0.1 respectively. And the losses are 1.464, 0.780, 0.774, 0.763, 0.761, respectively.



- The table which is giving information on the losses of the neural networks.

Losses on the Normalized Training Set		**Losses on the Normalized Test Set**
Hidden Unit = 2	1.465	1.284
Hidden Unit = 4	0.785	0.740
Hidden Unit = 8	0.783	0.766
Hidden Unit = 16	0.764	0.738
Hidden Unit = 32	0.761	0.737

As can be observed from the table above, it is not surprising that the model experiences less loss as the hidden unit count increases. As the model gets complicated by adding hidden units, its performance on this dataset increases which is proven by the decreasing loss values. Additionally, the neural networks do not overfit as the hidden unit count increases. We can infer this from the fact that the test set loss decreases. If it was overfitted, the training loss would get closer to zero, and the test loss would increase. Since we don’t observe this situation, overfitting is not happening in this model.