Necati Furkan Çolak / 21803512 / IE

Question 1: The initial parameters: Epoch:50, learning rate:0.001, initial weights and biases: 0.5.

a.1): Employing precisely two hidden units in an ANN can be sufficient to surpass linear regression's performance. Therefore, I choose my hidden unit number as 2.

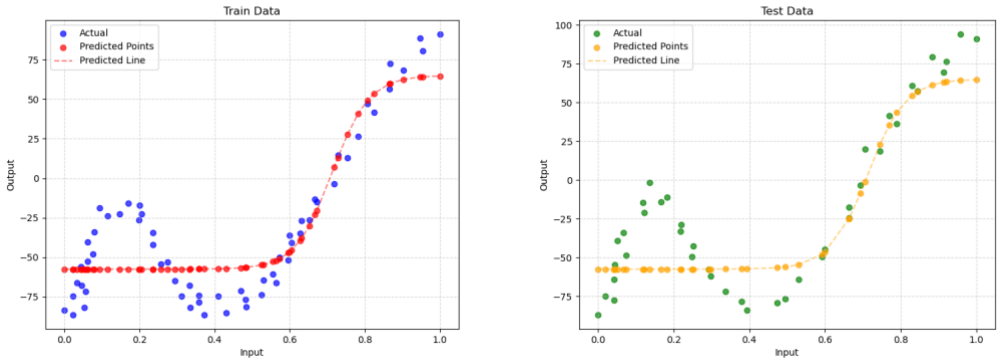
a.2): An exploration of diverse learning rates on model performance revealed intriguing findings. The learning rate of 0.5 caused swift convergence, impairing learning. A rate of 0.0000001 crippled the model's learning capabilities. A rate of 0.01 showed better results, while 0.001 achieved optimal performance. The findings highlight the crucial role of learning rate in model optimization. Careful rate selection is essential to avoid hindering the learning process. The rate of 0.001, providing the best results, was adopted for my configuration.

a.3): To examine the impact of weights on the model, I intentionally selected three different values: 0.2, 0.0001, and 0.5, while keeping the bias at 0.5. The model's performance improved slightly with a weight of 0.0001, but the change was negligible (R^2 changed from 0.7886 to 0.7885). However, using a weight of 0.5 resulted in a noticeable enhancement, with the R^2 value increasing to 0.7929. When a weight of 0.7 was used, the performance slightly decreased to around 0.7905. Based on these findings, I concluded that a weight of 0.5 produced the best outcomes for my model. Therefore, I decided to adopt this weight value as my initial weight, as it led to the highest R^2 value and improved performance compared to the other weights tested.

a.4): After evaluating our findings, we determined that training for 1500 epochs is the optimal choice. Extending training to 1500 epochs showed only marginal decreases in average loss beyond the 1200th epoch, resulting in no notable improvement in model performance. Concluding training at 700 epochs provides satisfactory results without significant gains beyond that point. While there is a slight improvement beyond 700 epochs, the performance of the model does not increase significantly.

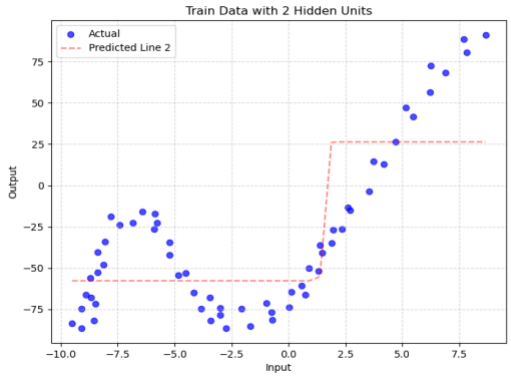
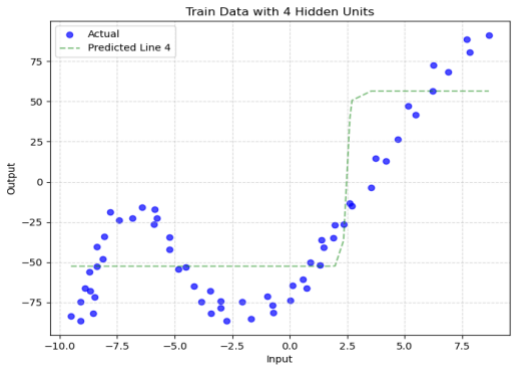
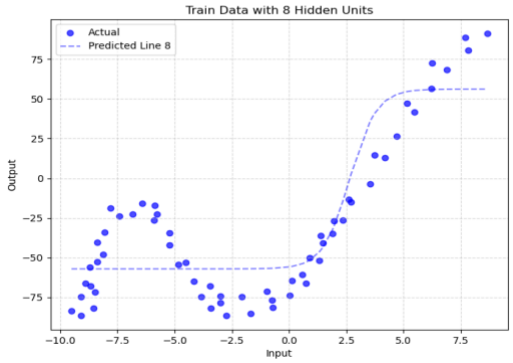
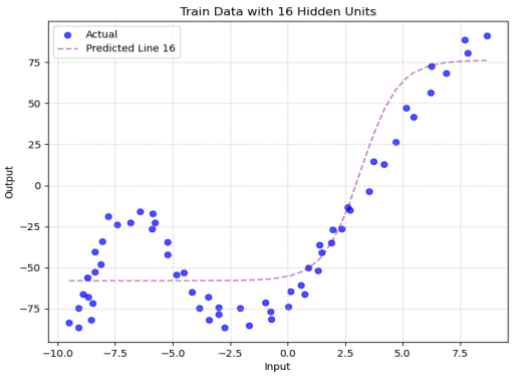
a.5): The implementation of normalization has proven to yield superior results in our case. Specifically, we have opted for min-max normalization. This normalization approach has led to notable enhancement in the coefficient of determination (R^2), increasing it from 0.79 to 0.83. This is a significant improvement in the model's performance.
Final configuration: Epoch:1000, learning rate:0.001, initial weights and biases: 0.9 and normalization applied.

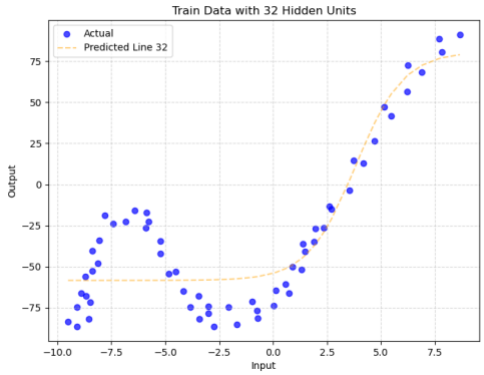
Question 2:



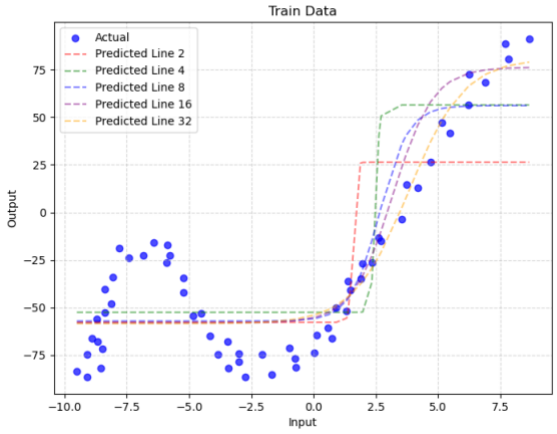
ANN used (specify the number of hidden units): ANN is applied with 2 hidden units.
Learning rate: 0.001
Range of initial weights: [-0.9,0.9]
Number of epochs: 1000
When to stop: 600
Is normalization used: Normalization is used.
Training loss (averaged over training instances): 360.90 for 901 epoch.
Test loss (averaged over test instances): 467.33 for 901 epochs.

Question 3.a)

ANN used (specify the number of hidden units): 2 **Learning rate:** 0.005 **Range of initial weights: [-0.5,0.5]** **Number of epochs:** 3000 **When to stop**: 1200 **Is normalization used:** Normalization is not used. **Training loss (averaged over training instances):** 612.67 **Test loss (averaged over test instances):** 709.93	
ANN used (specify the number of hidden units): 4 **Learning rate:** 0.004 **Range of initial weights: [-0.6,0.6]** **Number of epochs:** 3000 **When to stop**: 1200 **Is normalization used:** Normalization is not used. **Training loss (averaged over training instances):** 609.868. **Test loss (averaged over test instances):** 638.85	
ANN used (specify the number of hidden units): 8 **Learning rate:** 0.003 **Range of initial weights: [-0.7,0.7]** **Number of epochs:** 3000 **When to stop**: 1200 **Is normalization used:** Normalization is not used. **Training loss (averaged over training instances):** 499.65. **Test loss (averaged over test instances):** 539.04.	
ANN used (specify the number of hidden units): 16 **Learning rate:** 0.002 **Range of initial weights: [-0.8,0.8]** **Number of epochs:** 1000 **When to stop**: 201 **Is normalization used:** Normalization is not used. **Training loss (averaged over training instances):** 376.57 **Test loss (averaged over test instances):** 502.00.	

ANN used (specify the number of hidden units): 32 **Learning rate:** 0.001 **Range of initial weights:** [-0.9,0.9] **Number of epochs:** 1000 **When to stop**: 201 **Is normalization used:** Normalization is not used. **Training loss (averaged over training instances):** 353.93 for 501 epoch. **Test loss (averaged over test instances):** 454.44 for 501 epoch.	

General Plot:



Question 3.b):

Parameters: 2 hidden units, learning rate: 0.005, initial weight range: [-0.5,0.5]

Epoch Numbers	Average Train Loss	Average Test Loss
1	3136.62	2888.35
101	838.12	1014.48
201	837.38	1013.05
301	836.05	1011.70
401	836.56	976.13
501	833.67	1007.97

Standard Deviation for Train Data: 24.65 / Standard Deviation for Test Data: 25.95

Parameters: 4 hidden units, learning rate: 0.004, initial weight range: [-0.6,0.6]

Epoch Numbers	Average Train Loss	Average Test Loss
1	3186.60	2934.26
101	806.74	889.67
201	624.22	823.68
301	640.00	667.33
401	602.99	707.20
501	626.36	684.56

Standard Deviation for Train Data: 24.22 / Standard Deviation for Test Data: 25.58

Parameters: 8 hidden units, learning rate: 0.003, initial weight range: [-0.7,0.7]

Epoch Numbers	Average Train Loss	Average Test Loss
1	3238.77	3000.17
101	494.97	594.59
201	505.65	660.72
301	452.51	569.12
401	422.81	620.39
501	438.51	556.42

Standard Deviation for Train Data: 20.08 / Standard Deviation for Test Data: 21.56

Parameters: 16 hidden units, learning rate: 0.002, initial weight range: [-0.8,0.8]

Epoch Numbers	Average Train Loss	Average Test Loss
1	3309.82	3059.05
101	376.55	501.78
201	376.57	502.00
301	376.57	502.00
401	376.57	502.00
501	376.57	502.00

Standard Deviation for Train Data: 20.06 / Standard Deviation for Test Data: 22.36

Parameters: 32 hidden units, learning rate: 0.001, initial weight range: [-0.9,0.9]

Epoch Numbers	Average Train Loss	Average Test Loss
1	3527.73	3021.12
101	356.02	443.03
201	353.97	453.47
301	353.94	454.35
401	353.93	454.43
501	353.93	454.44

Standard Deviation for Train Data: 18.82 / Standard Deviation for Test Data: 20.74

Question 3.c):

A minimal number of hidden units in an ANN could prevent it from learning complicated patterns in the data, which would result in underfitting. The findings suggest that as the artificial neural networks (ANN) hidden unit count rises, so does the model's capacity to absorb new information. As the number of concealed units increases from 2 to 32, this is reflected in the steadily declining Average Train Loss and Average Test Loss. Additionally, as hidden units are added, the standard deviation of both train and test data decreases, indicating a potential reduction in data variability as well as overfitting, particularly at hidden unit counts of 16 and 32, where the train loss and test loss begin to plateau. The findings show that an artificial neural network's ability to recognize intricate patterns in data is enhanced by adding more hidden units. As the number of hidden units increases from 2 to 32, the average train loss and average test loss both steadily decrease. However, care should be taken because overfitting is a problem when the train loss and test loss begin to plateau, especially with 16 and 32 hidden units, as seen by a decrease in data variability. Also as hidden unit number increases predicted plots become more curved since model becomes to capture more variance of the given points.