Mustafa Cankan BALCI

22101761

27 April 2025

GE 461 – Introduction to Data Science Project 3

The project aims to create and observe a neural network structurer from the scratch. The training process of the model is declared in the assignment document. The loss of neural network is sum of squared errors. The activation function in the hidden layer is sigmoid. The input of ANN is 1 dimensional, and outputs of ANN is 1 dimensional.

Part A

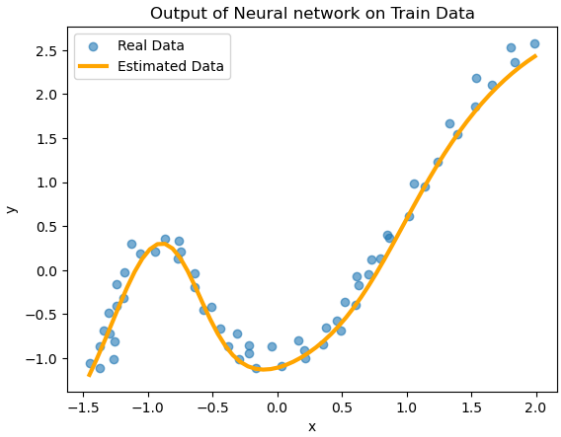
In the assignment, I tried different hyperparameters and approaches for finding the configuration that learns a network on its training instances. First of all, linear regression and single layer ANN is compared. The data points have high order polynomial shape, so they have nonlinear relationship. As a result, a single layer ANN performed better than linear regressor in this task. Also, the obtaining more information about the data, the hidden layer number has to be bigger than 1. There is no specific parameter that this learning rate is better, but it is chosen as 0.075 which is lower than 0. The weights are initialized randomly between - 0.01 and 0.01 since neural network can break symmetry and converge effectively without saturating. The epoch number is selected between 300 and 1000 according to convergence speed and stability. Early stooping is applied to overcome the issue of overfitting. The normalization effects the performance of model and stability, so normalizing inputs and outputs facilitated smoother convergence and prevented saturation issues.

Part B

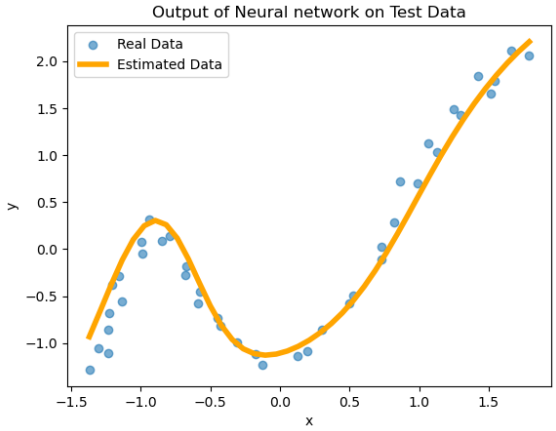
The single hidden layer neural network structure is designed in part B. The parameters of the model in the training process and outputs are shown below. Table 1 represents details about parameters and result of training process.

ANN used (specify the number of hidden units)	8
Learning Rate	0.075
Range of Initial Weights	[-0.01, 0,01]
Number of Epochs	1000
When to Stop	58
Is normalization used	Yes
Training Loss (averaged over test instances)	0.032
Test Loss (averaged over test instances)	0.050

Table 1: Part B Results



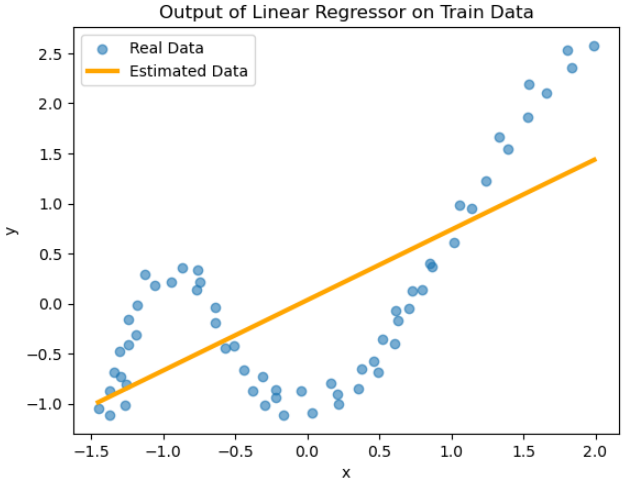
Graph 1: Output of ANN on Train Data Part B



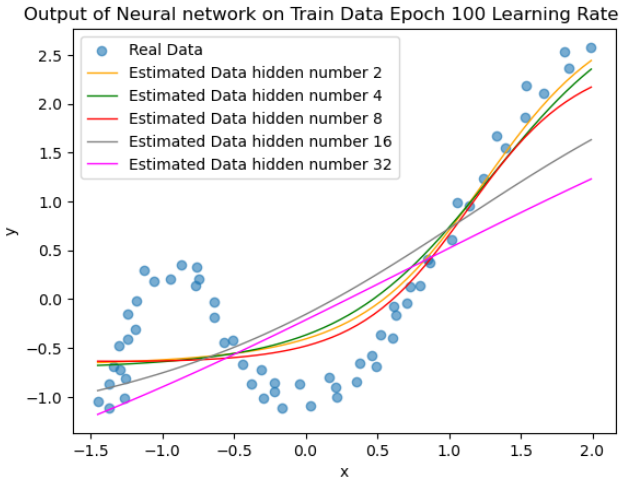
Graph 2: Output of ANN on Test Data Part B

Part C

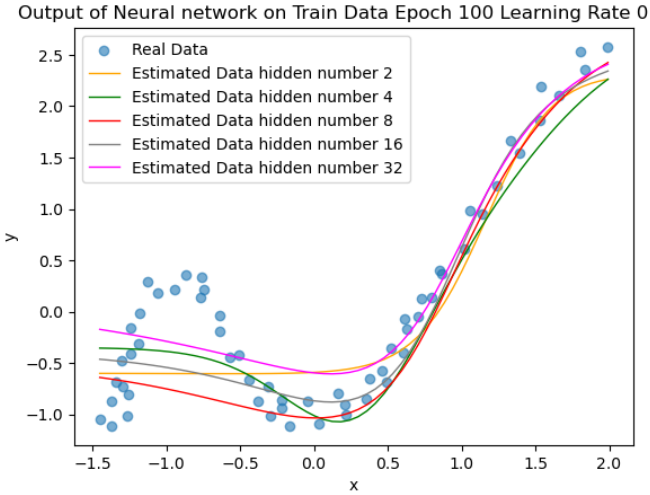
In this part of assignment, different hidden layer number is tried initially. 2,4,8,16 and 32 hidden units are tried. For each set, different learning rates and initial weights also tried.



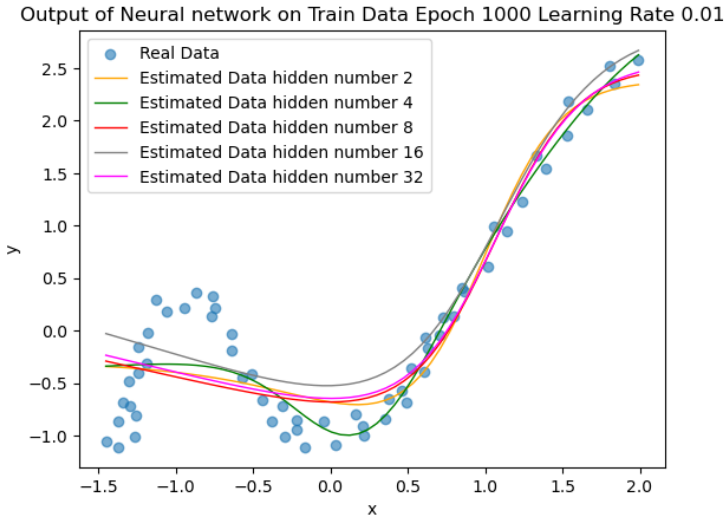
Graph 3: Output of Linear Regressor on Train Data



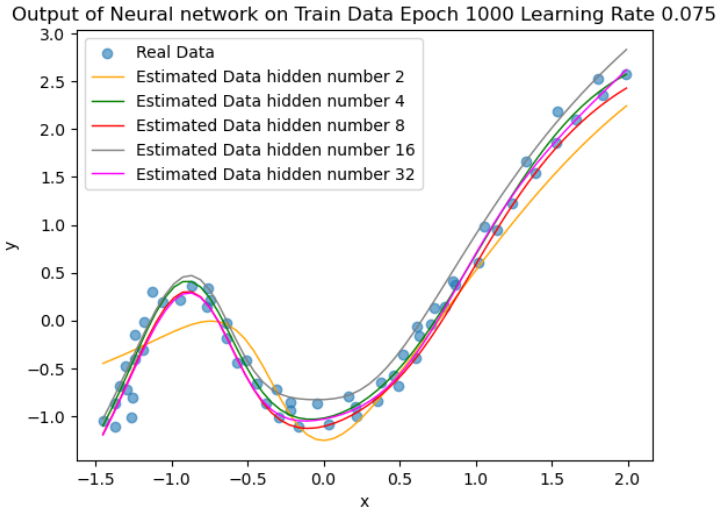
Graph 4: Output of ANN on Train Data when Epoch 100 Learning Rate 0.01



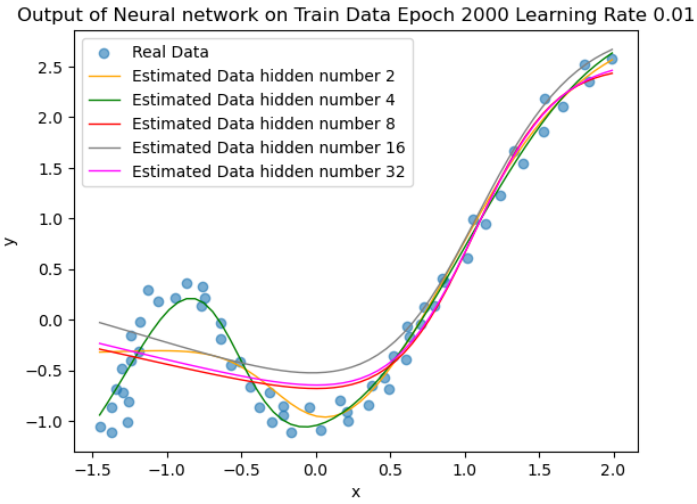
Graph 5: Output of ANN on Train Data when Epoch 100 Learning Rate 0.075



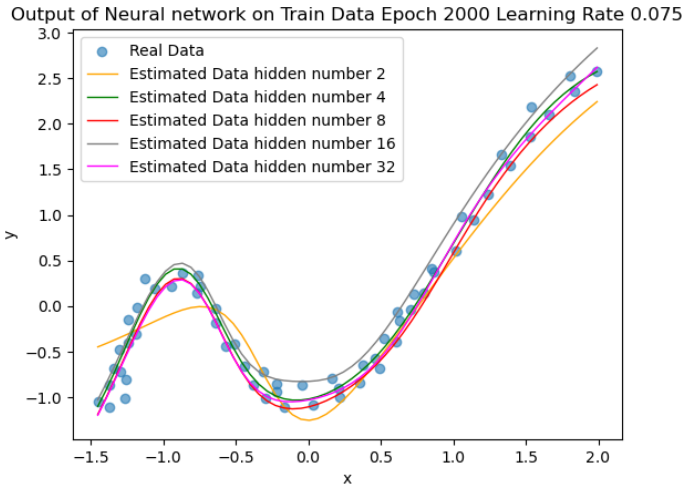
Graph 6: Output of ANN on Train Data when Epoch 1000 Learning Rate 0.01



Graph 7: Output of ANN on Train Data when Epoch 1000 Learning Rate 0.075



Graph 8: Output of ANN on Train Data when Epoch 2000 Learning Rate 0.01



Graph 9: Output of ANN on Train Data when Epoch 2000 Learning Rate 0.075

Number of Epochs	Learning Rate	Number of hidden layer	Train Loss Average	Train Loss Standard Deviation	Test Loss Average	Test Loss Standard Deviation
100	0.01	2	0.206	0.240	0.197	0.223
100	0.01	4	0.234	0.252	0.216	0.240
100	0.01	16	0.201	0.239	0.187	0.209
100	0.01	8	0.422	0.360	0.355	0.341
100	0.01	32	0.537	0.518	0.455	0.392
100	0.075	2	0.165	0.224	0.172	0.194
100	0.075	4	0.138	0.164	0.135	0.183
100	0.075	16	0.206	0.344	0.180	0.274

100	0.075	8	0.141	0.220	0.133	0.190
100	0.075	32	0.141	0.192	0.133	0.246
1000	0.01	2	0.129	0.168	0.142	0.188
1000	0.01	4	0.098	0.153	0.110	0.186
1000	0.01	8	0.138	0.195	0.152	0.202
1000	0.01	16	0.175	0.237	0.209	0.302
1000	0.01	32	0.140	0.192	0.158	0.215
1000	0.075	2	0.093	0.117	0.121	0.172
1000	0.075	4	0.021	0.041	0.071	0.102
1000	0.075	8	0.032	0.042	0.050	0.082
1000	0.075	16	0.040	0.054	0.108	0.112
1000	0.075	32	0.025	0.038	0.042	0.071
2000	0.01	2	0.098	0.154	0.113	0.191
2000	0.01	4	0.027	0.047	0.061	0.089
2000	0.01	8	0.138	0.195	0.152	0.202
2000	0.01	16	0.175	0.237	0.209	0.302
2000	0.01	32	0.140	0.192	0.158	0.215
2000	0.075	2	0.093	0.117	0.121	0.172
2000	0.075	4	0.021	0.041	0.071	0.102
2000	0.075	8	0.032	0.042	0.050	0.082
2000	0.075	16	0.040	0.054	0.108	0.112
2000	0.075	32	0.025	0.038	0.042	0.071

Table 2: Part C Results

When the number of hidden units increases, the complexity of ANN will increase. This makes the model more powerful in terms of learning detailed patterns from the training data. In theory, it helps the network understand the data better. However, model starts to memorizing the the training data. This issue is called overfitting since it learns more specific information about training set rather than general information of training set. In this assignment, the hidden size number 2 is underfits but the other parameters fit the data without underfitting. However, it might be overfit in 42 hidden layer.