

CS6910: Fundamentals of Deep Learning
Team 6 - Assignment 1 Report

Lavanya L (EE18D414)

Chigullapally Sriharsha (CS17D012)

Sivaraman S (OE19S012)

- 1. Function approximation on 2-d input**
- 2. Classification on 2d Non-linear data**
- 3. Image Classification**

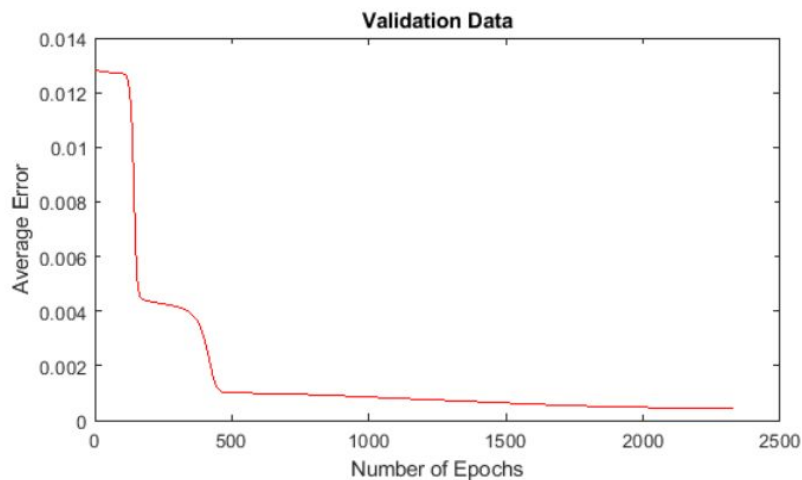
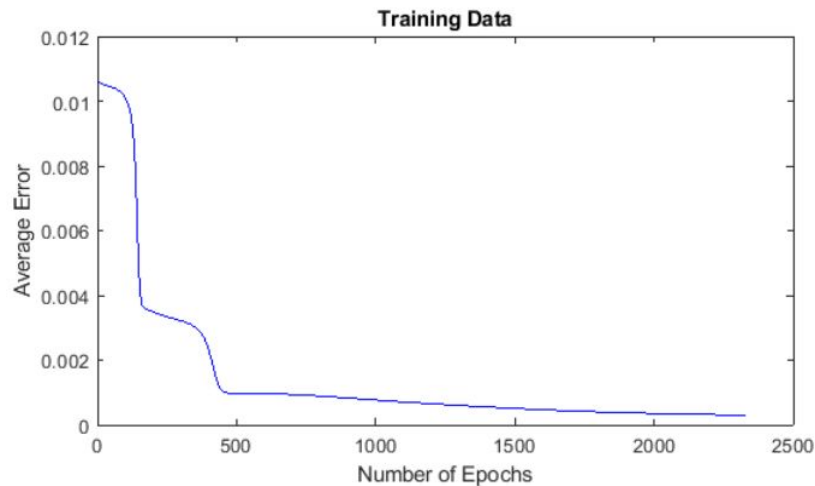
1. Function approximation on 2-d input

(1) Average Error of Validation data calculated using the final weights after training

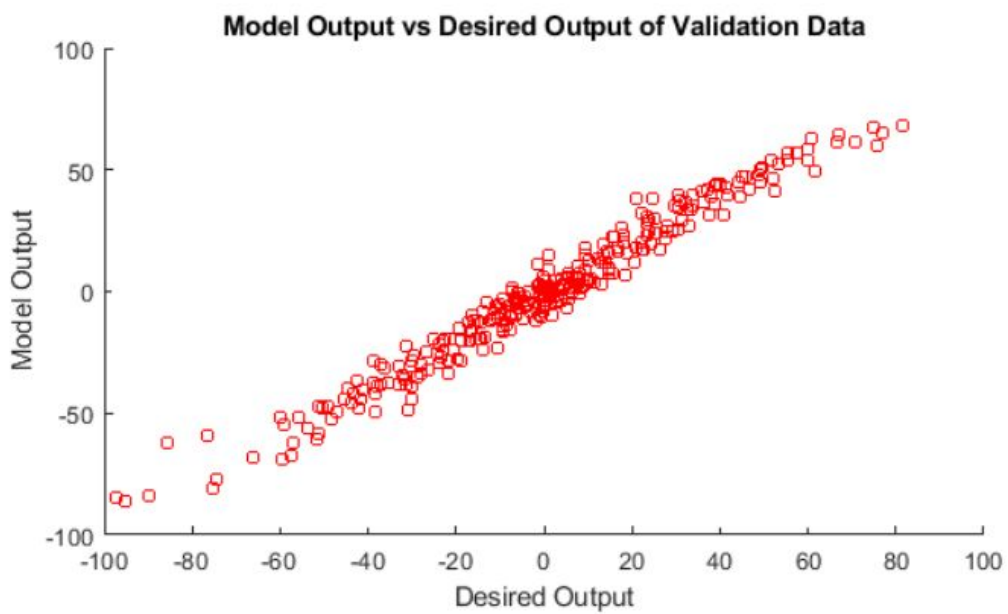
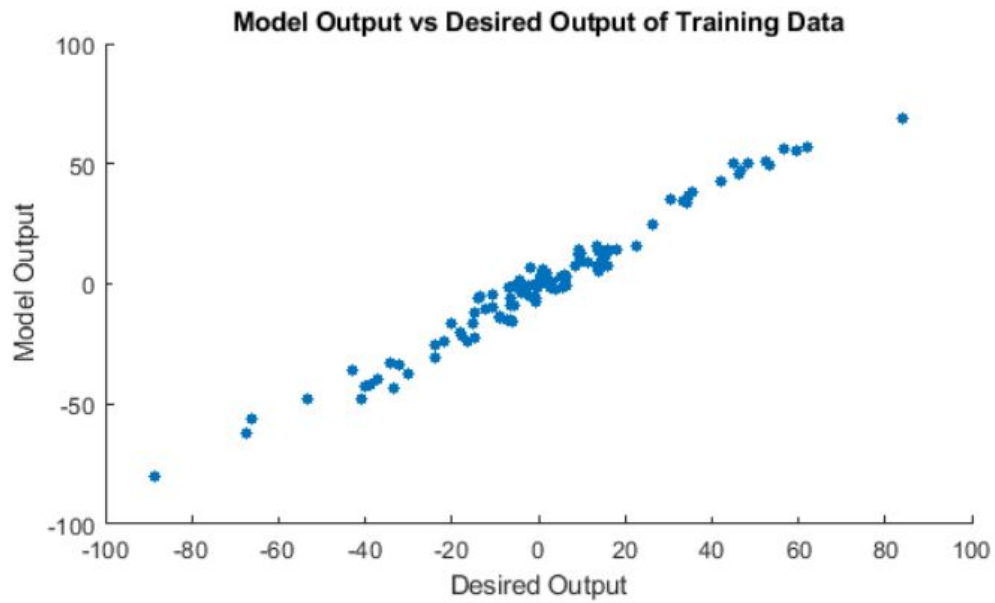
		Number of Nodes in the Hidden Layer 2			
Number of Nodes In the Hidden Layer 1	# Nodes	2	4	6	8
	2	0.0065	0.0069	0.0032	0.0078
	4	0.0016	0.0031	0.0026	0.0073
	6	0.0029	0.0018	0.0023	0.0021
	8	0.0027	0.0019	0.0035	0.0026

Based on the above table, Number of nodes for Hidden Layer 1 and 2 are chosen as 4 and 2.

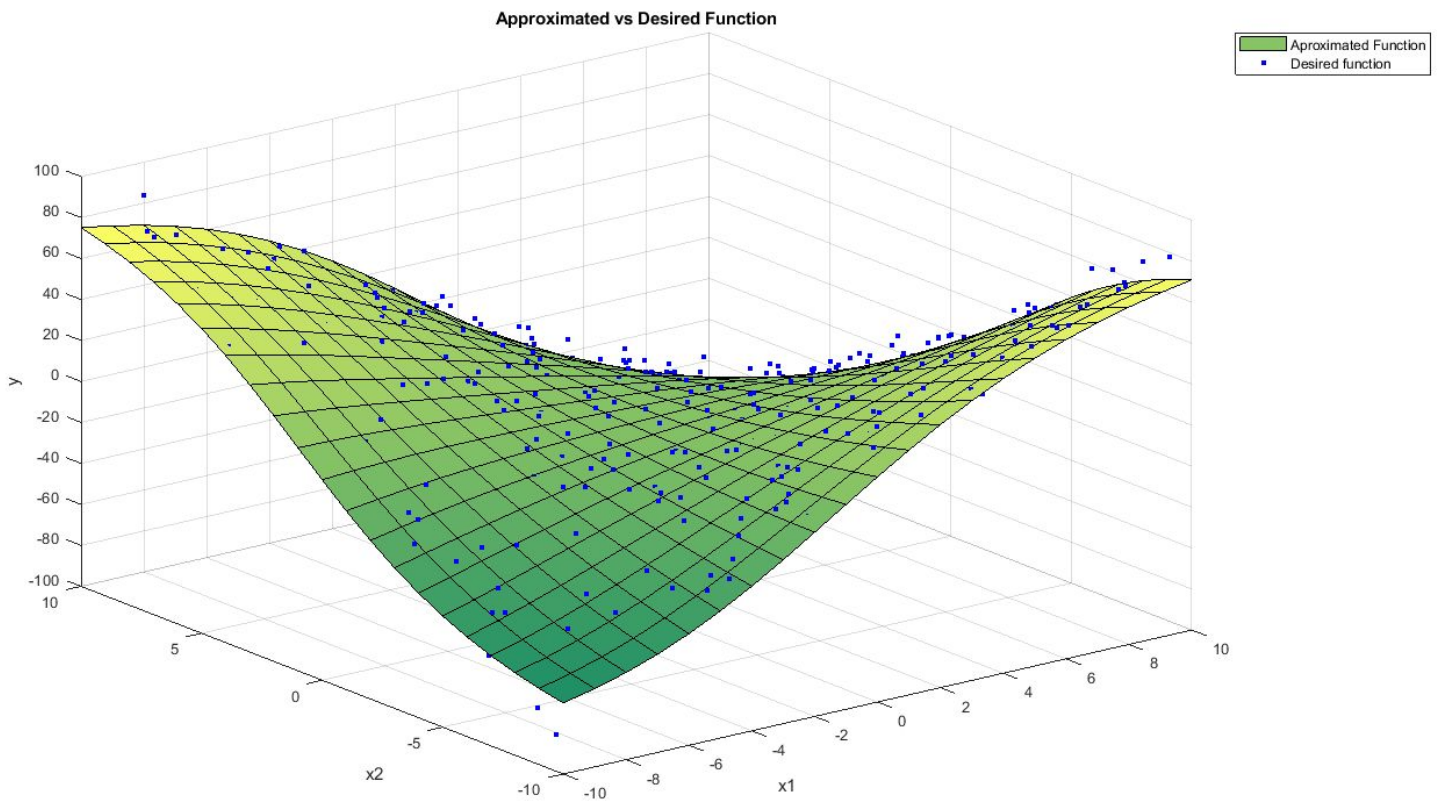
(2) Plot of Average error vs Epoch



(3) Scatter plot (Model output vs Desired output) after the model is trained



(4) Plots of the Desired vs Approximated functions



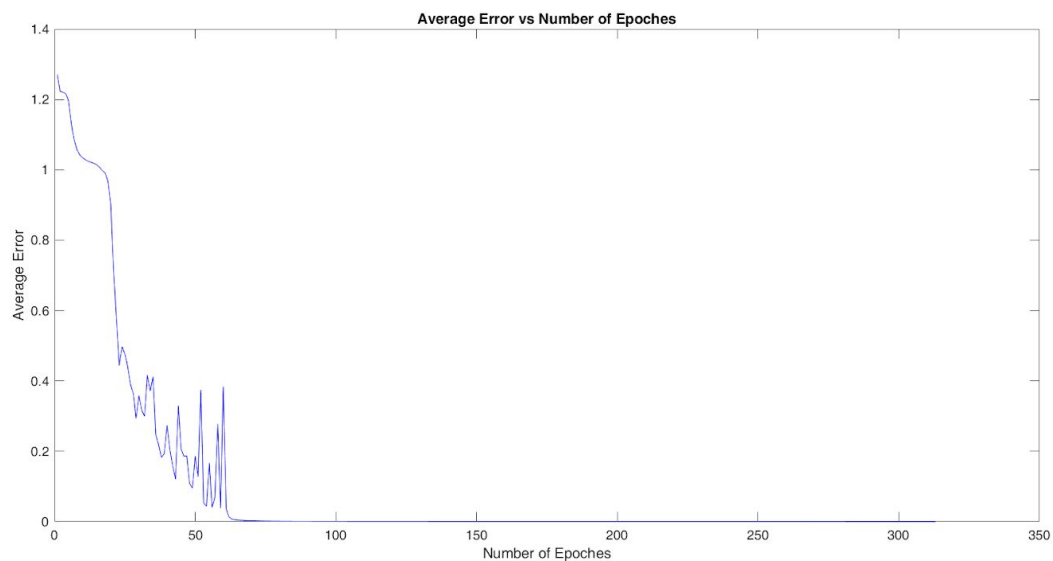
2. Classification on 2d Non-linear data

(1) Average Error of Validation data calculated using the final weights after training

		Number of Nodes in the Hidden Layer 2			
Number of Nodes In the Hidden Layer 1	# Nodes	4	6	8	10
	4	0.0194	0.0024	0.0035	0.0027
	6	0.0037	0.0020	0.0044	0.0043
	8	0.0073	0.0022	0.0121	0.0052
	10	0.0039	0.0059	0.0036	0.0033

Based on the above table, Number of nodes for Hidden Layer 1 and 2 are chosen as 6 and 6.

(2) Plot of Average error vs Epoch - Training Data



(3) Confusion Matrix:

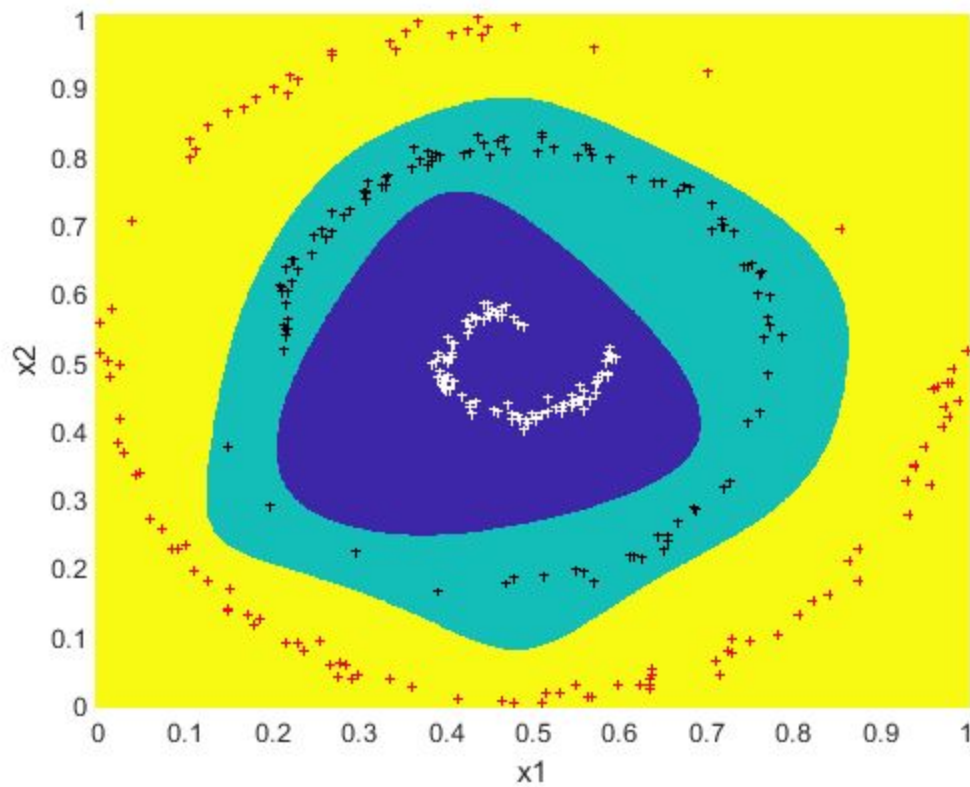
Training Data

98	0	0
0	108	0
0	0	109

Validation Data

53	0	0
0	40	2
0	0	41

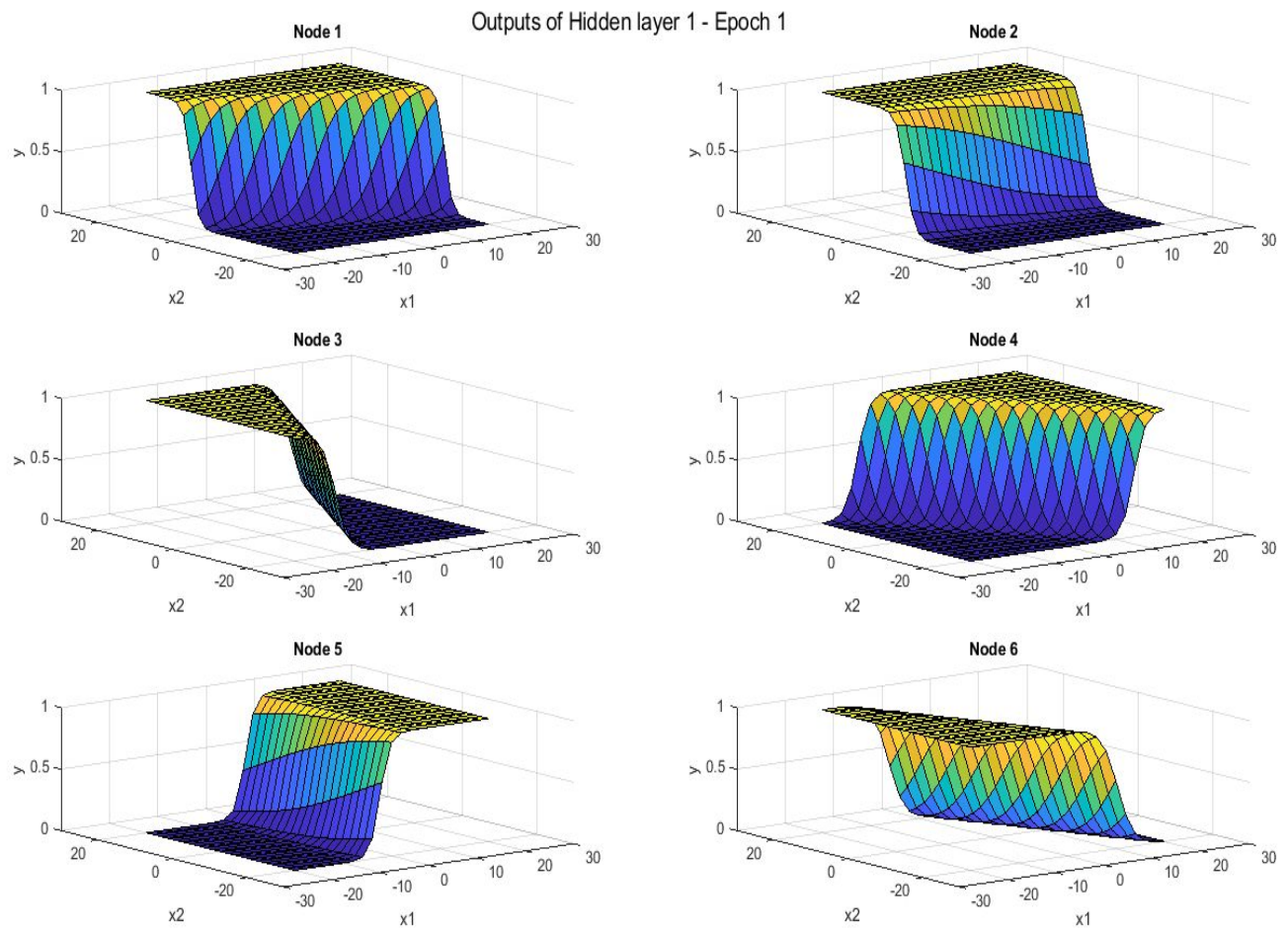
(4) Decision Boundary with Training Points superimposed



Blue region - class 0	White points - Training examples of class 0
Green region - class 1	Black points - Training examples of class 1
Yellow region - class 2	Red points - Training examples of class 2

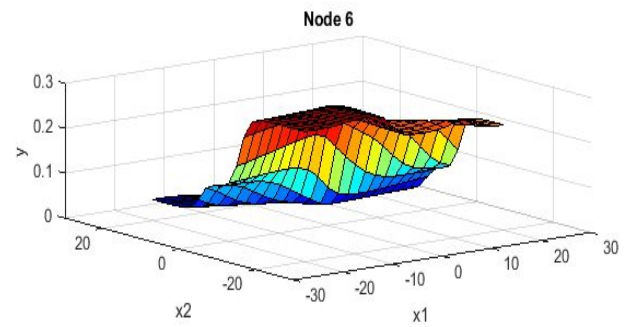
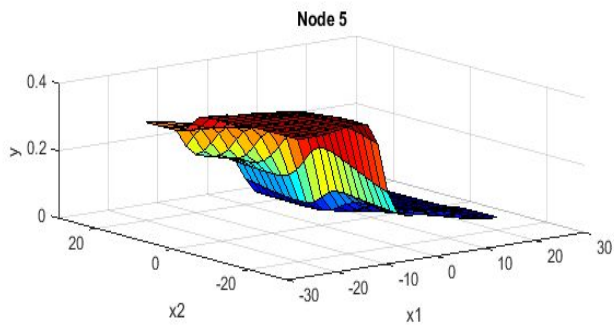
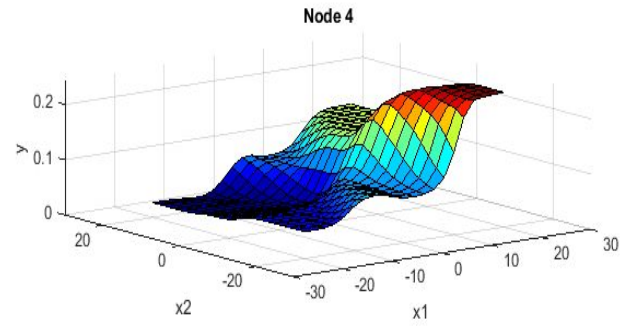
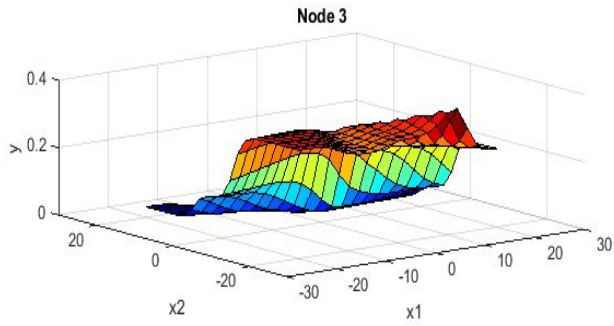
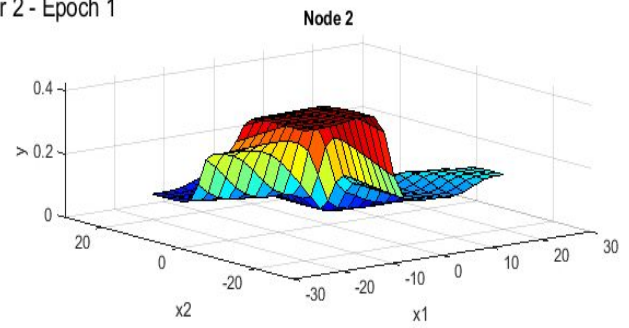
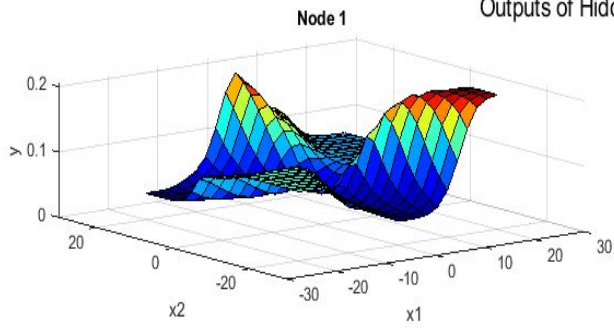
(5) Surfaces of outputs of nodes in hidden layers and output layer after Epoch 1

Hidden Layer 1:



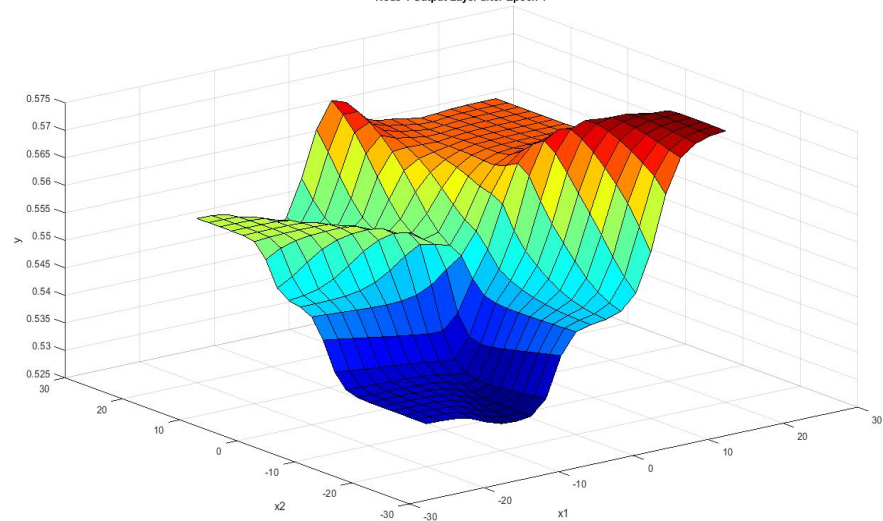
Hidden Layer 2

Outputs of Hidden layer 2 - Epoch 1

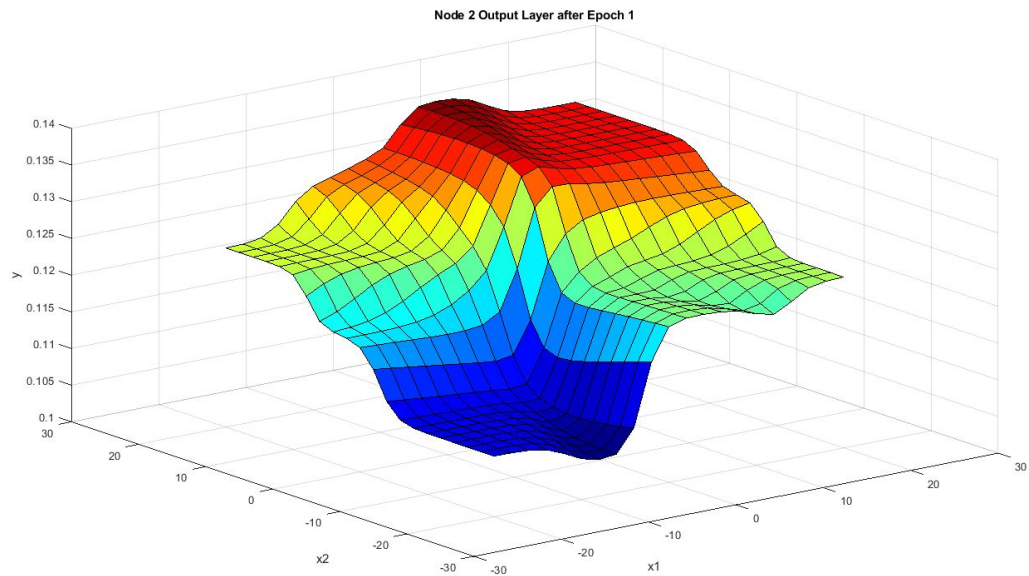


Node 1 of Output Layer

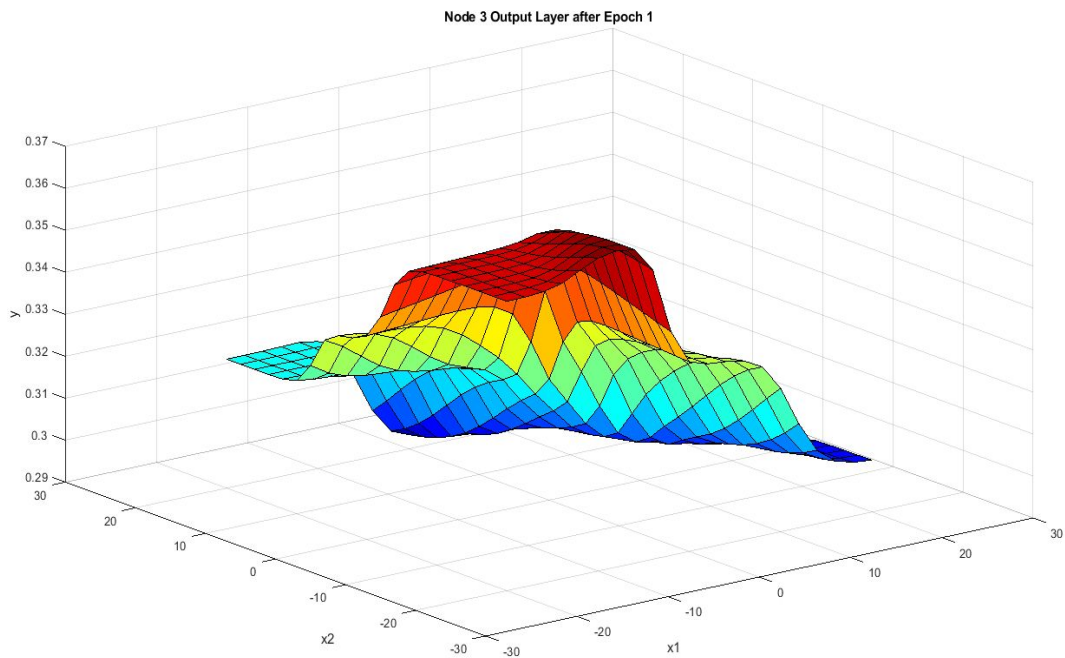
Node 1 Output Layer after Epoch 1



Node 2 of output layer

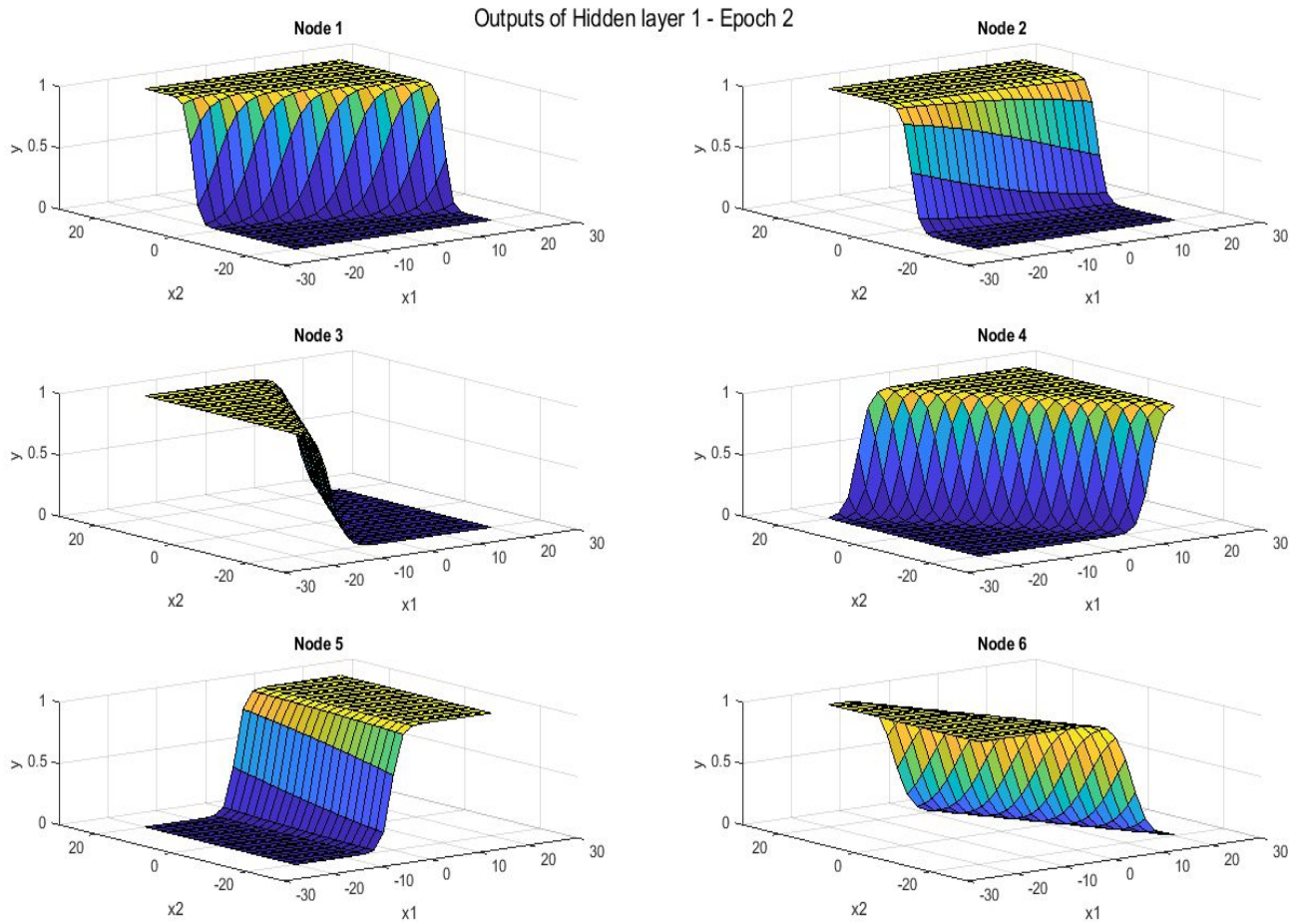


Node 3 of output layer

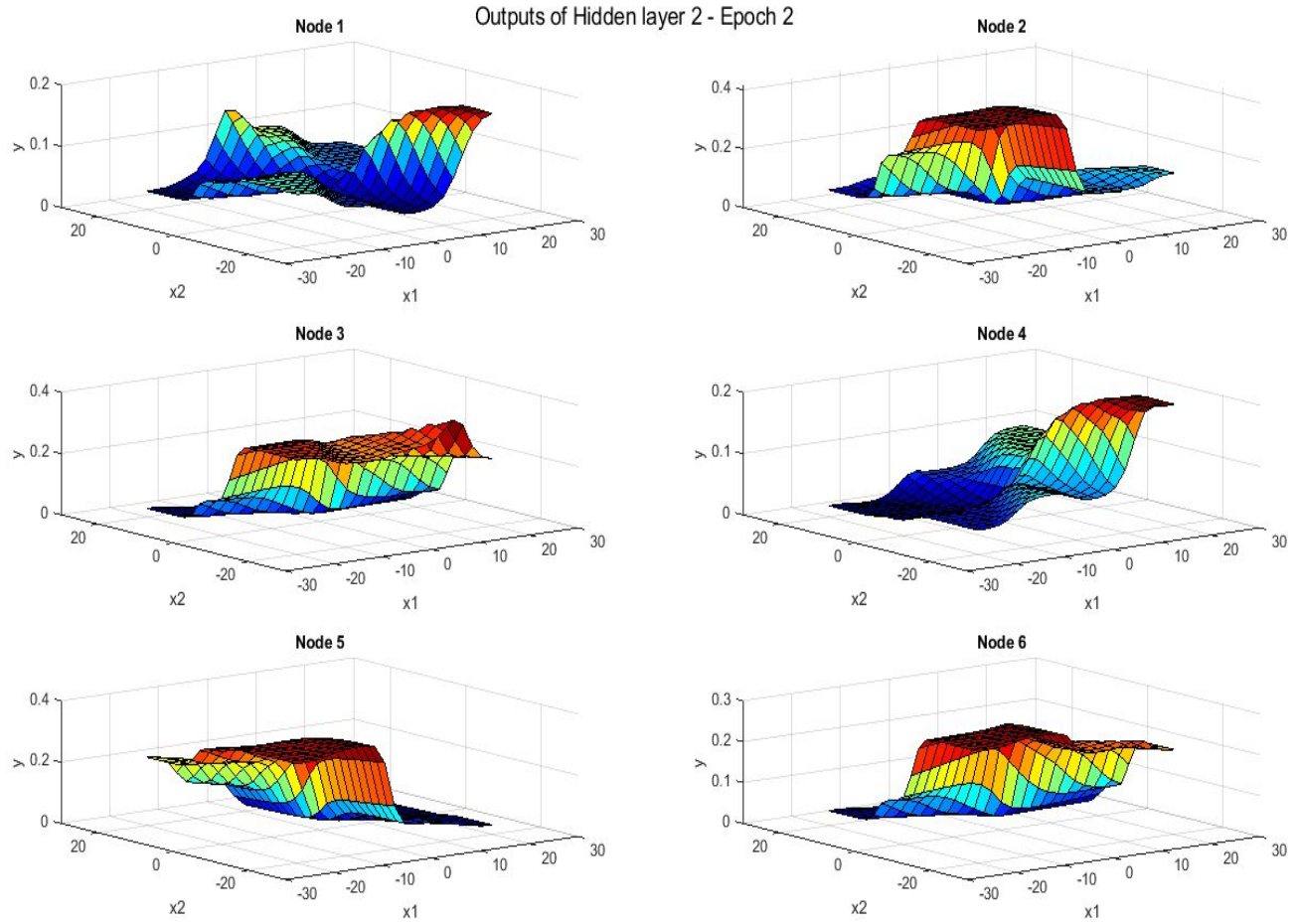


(6) Surfaces of outputs of nodes in hidden layers and output layer after Epoch 2

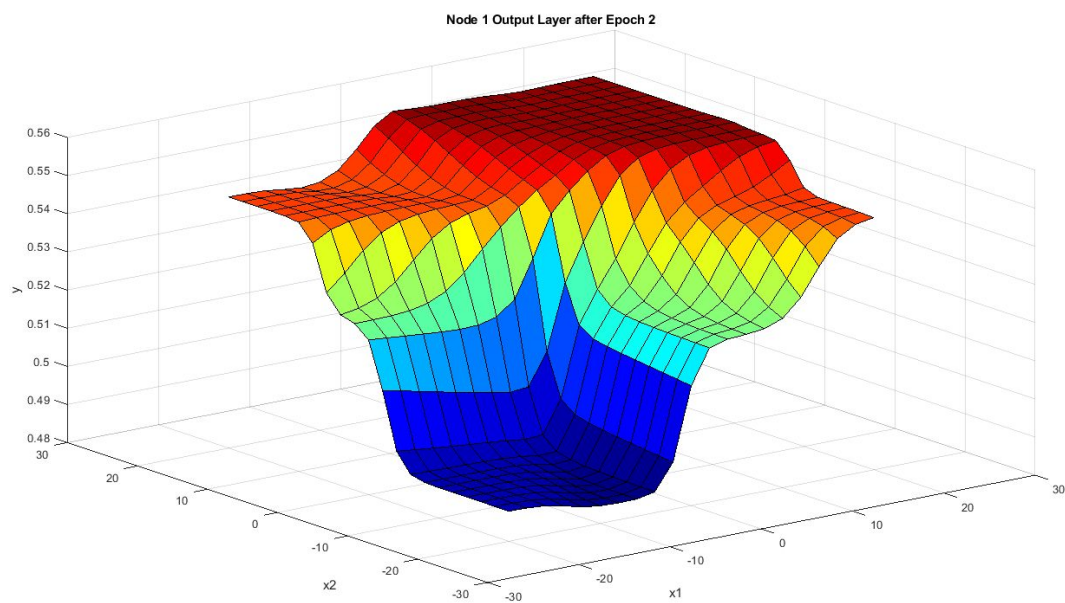
Hidden layer 1



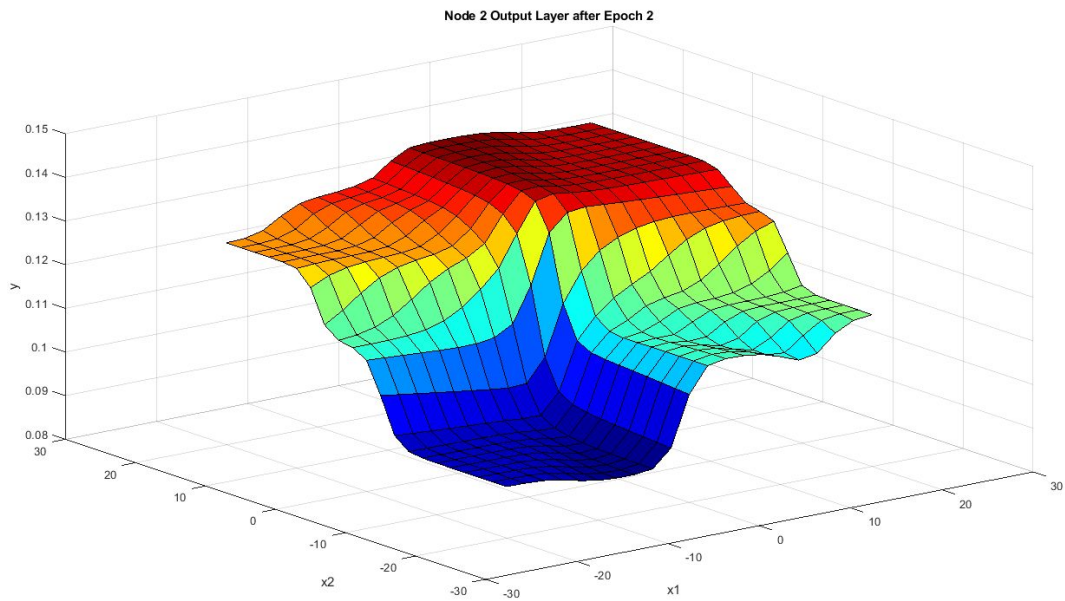
Hidden layer 2



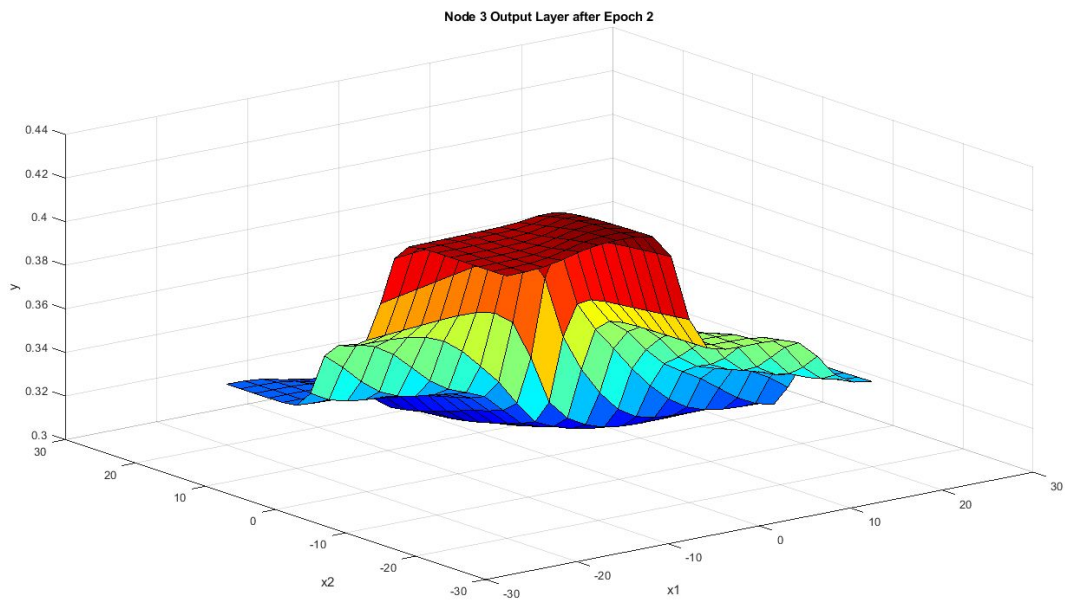
Node 1 of output layer



Node 2 of output layer

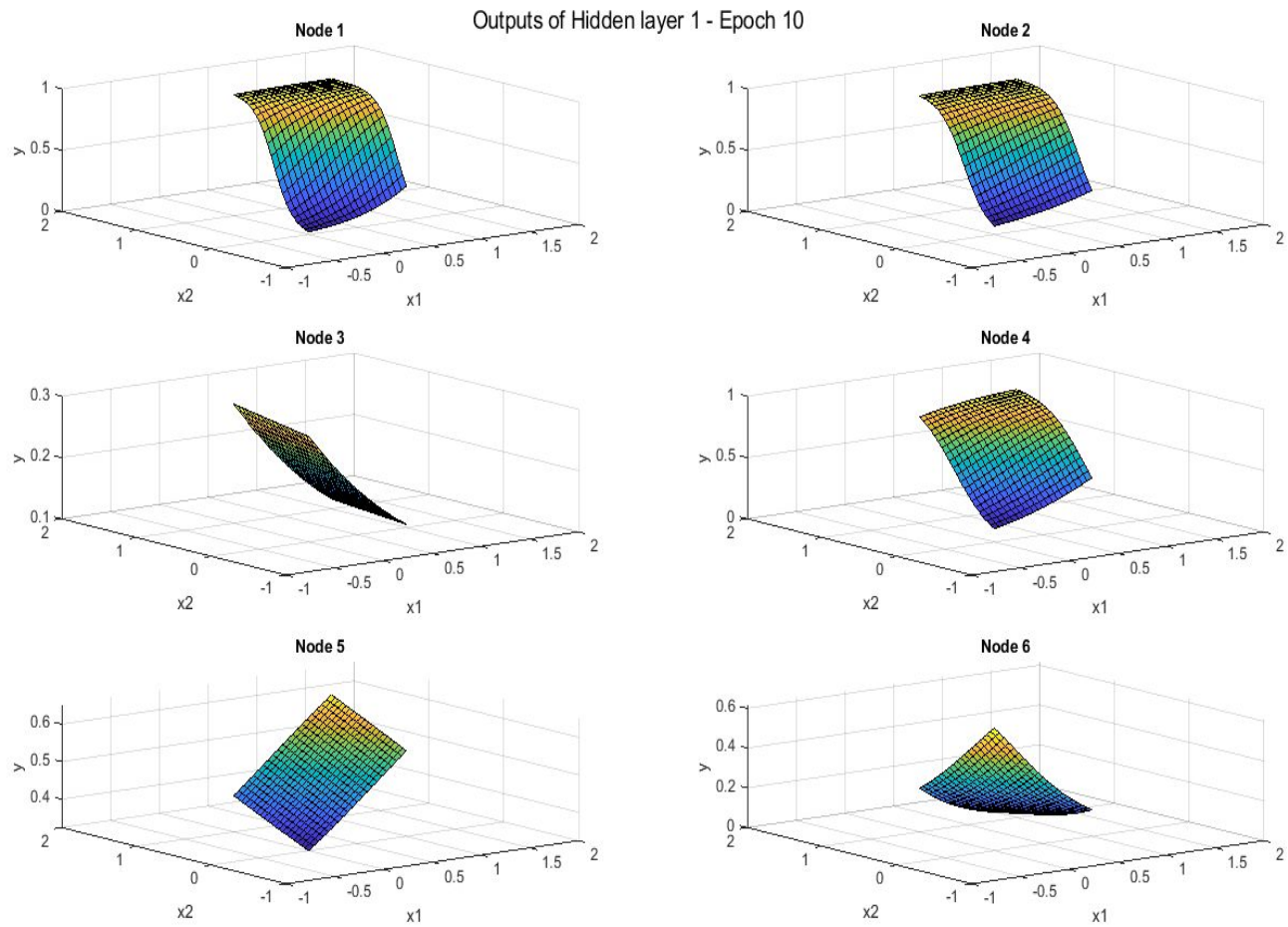


Node 3 of output layer

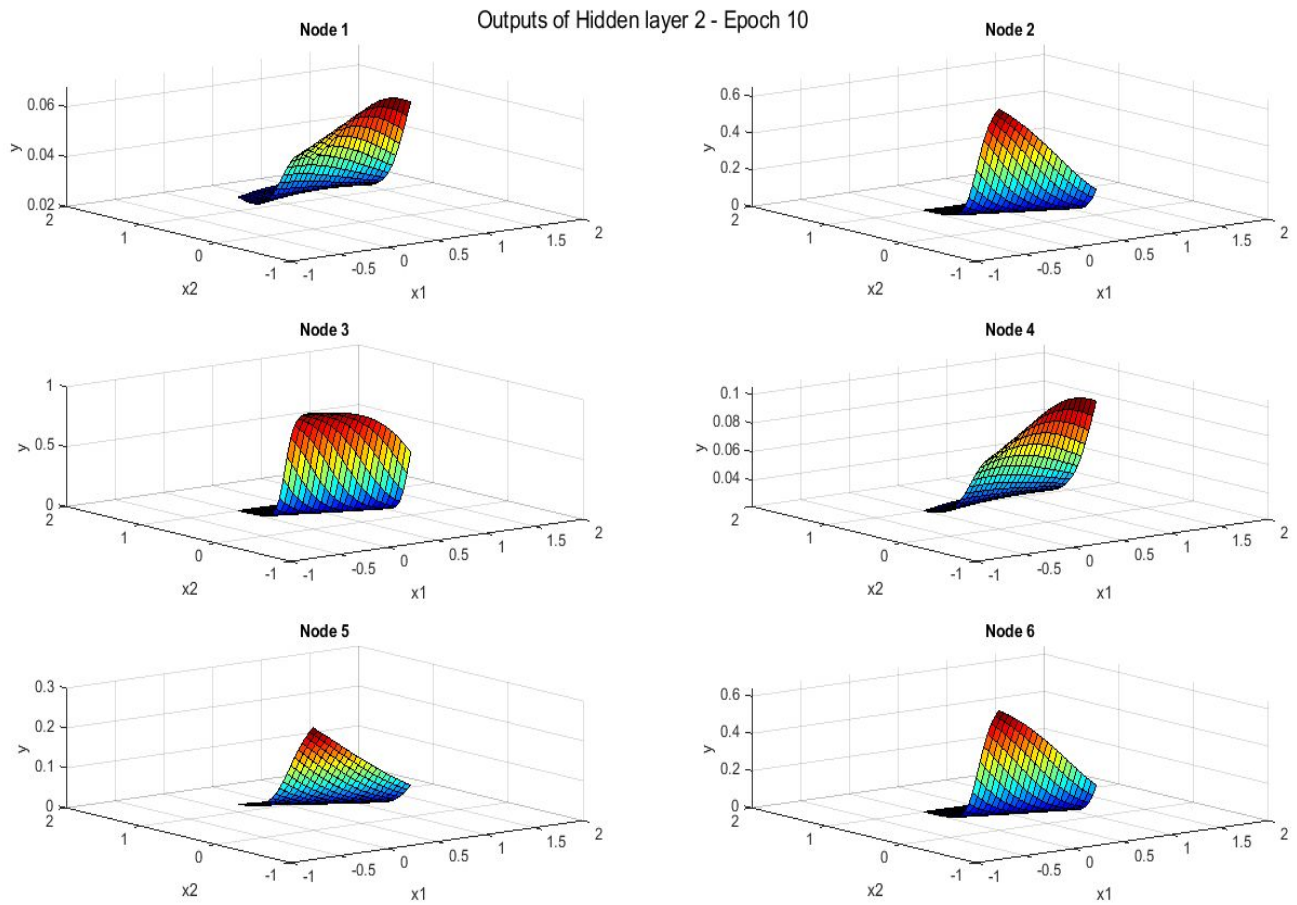


(7) Surfaces of outputs of nodes in hidden layers and output layer after Epoch 10

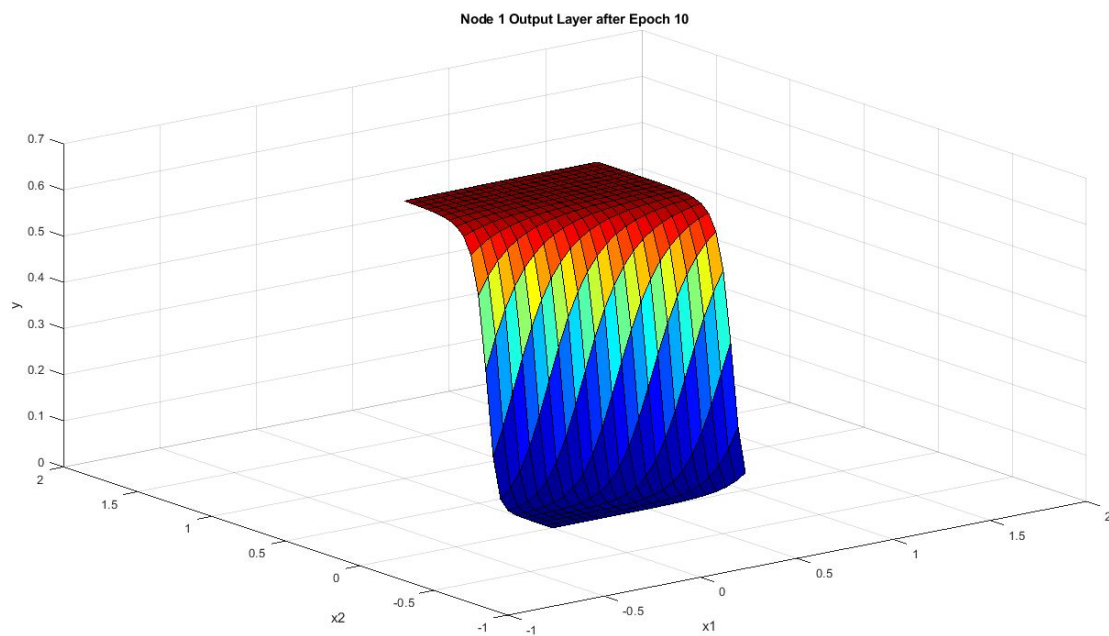
Hidden layer 1



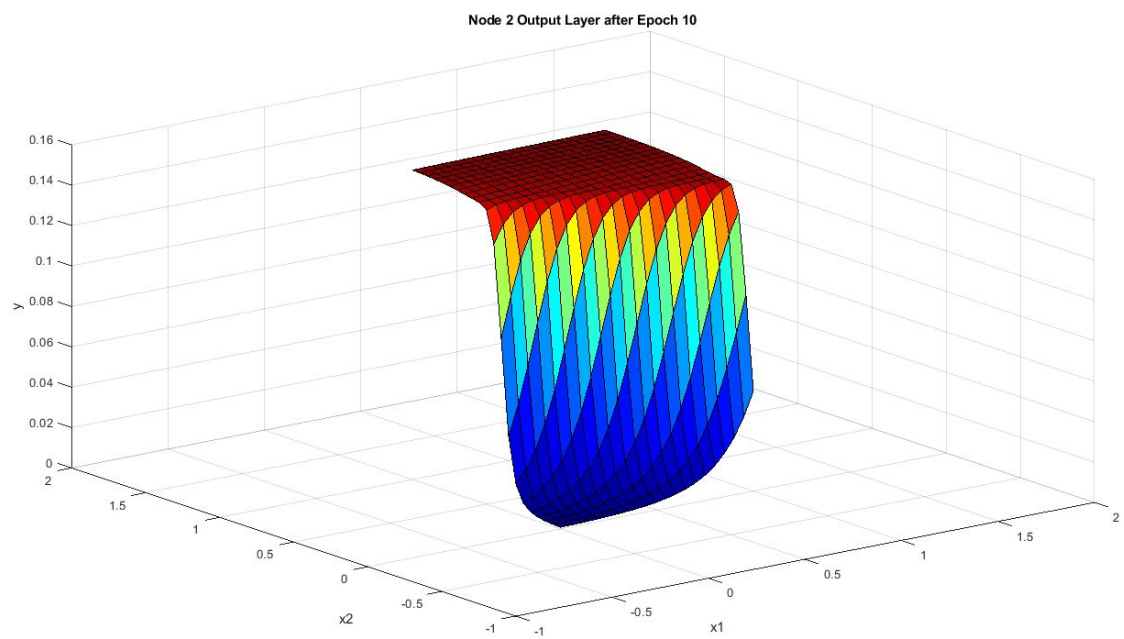
Hidden layer 2



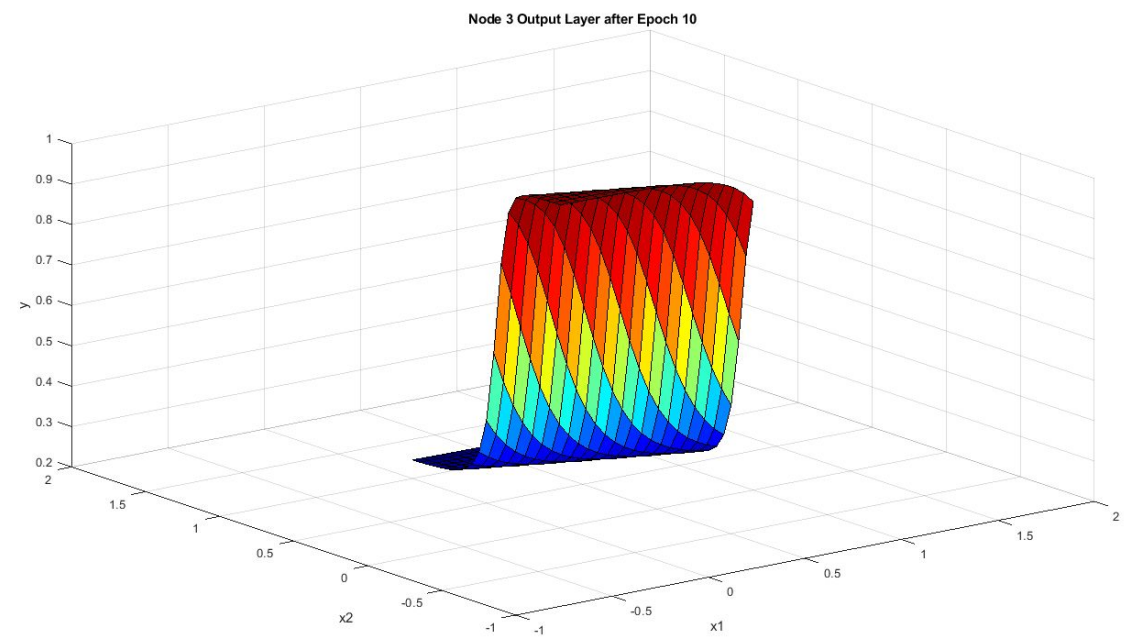
Node 1 of output layer



Node 2 of output layer

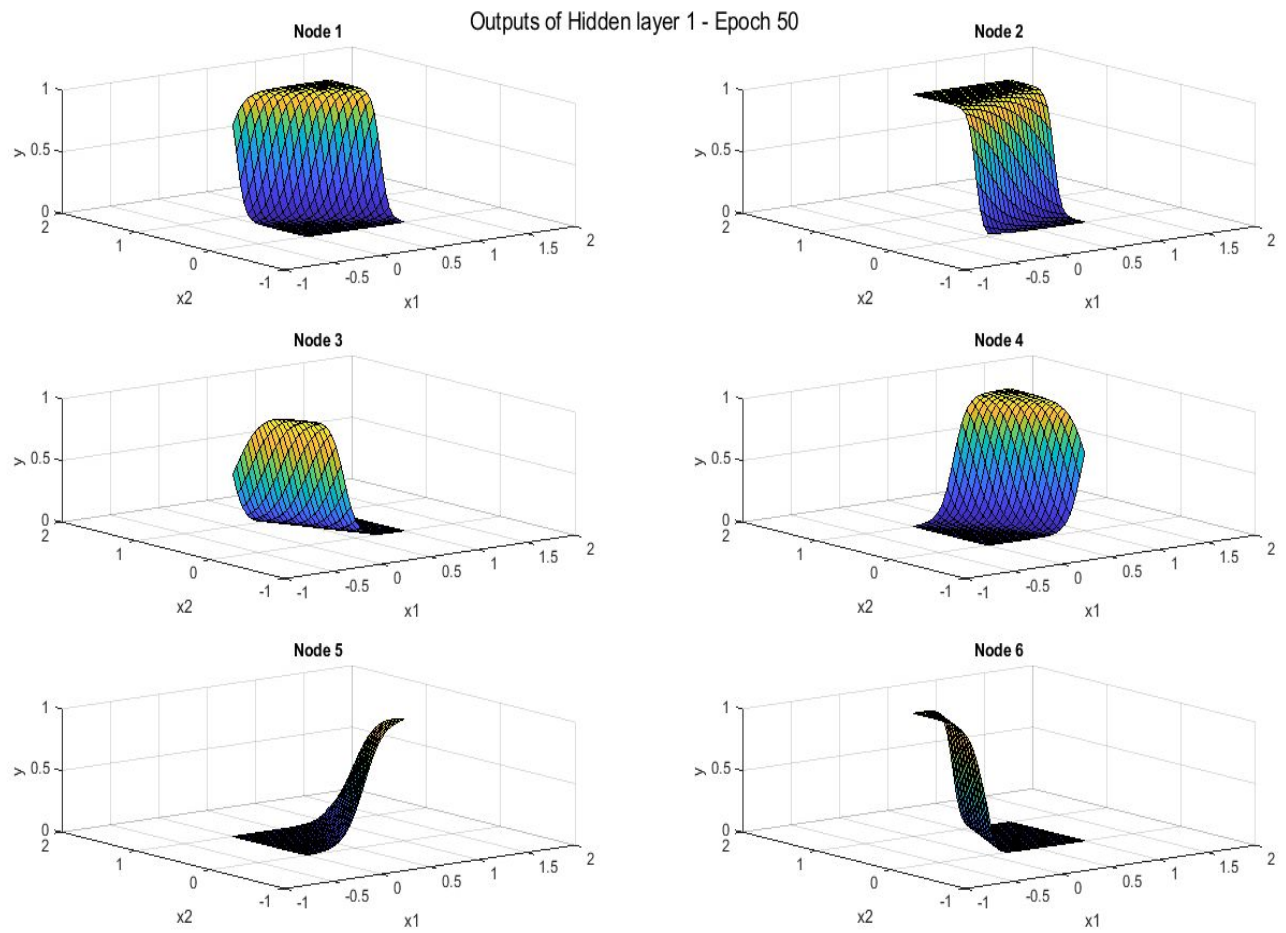


Node 3 of output layer

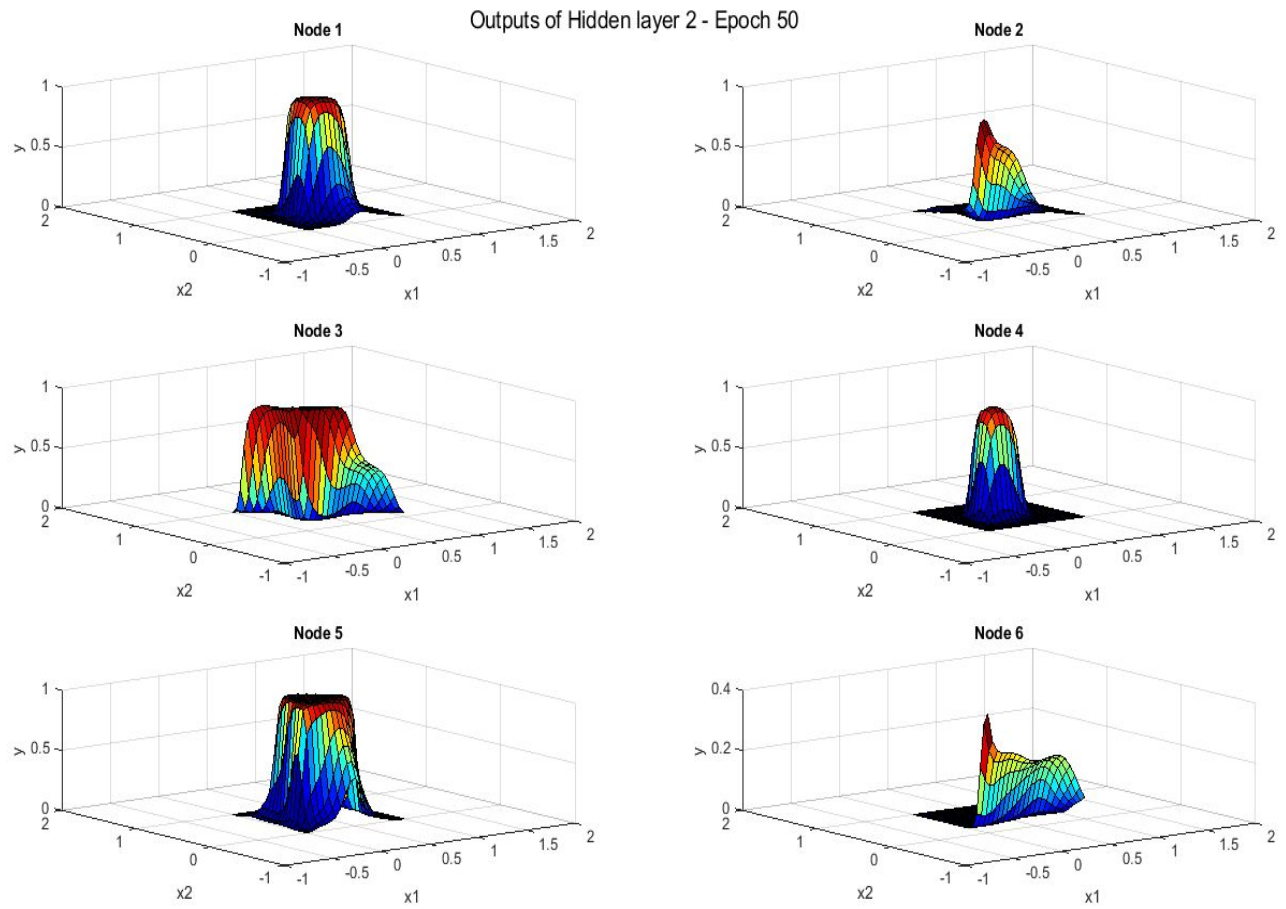


(8) Surfaces of outputs of nodes in hidden layers and output layer after Epoch 50

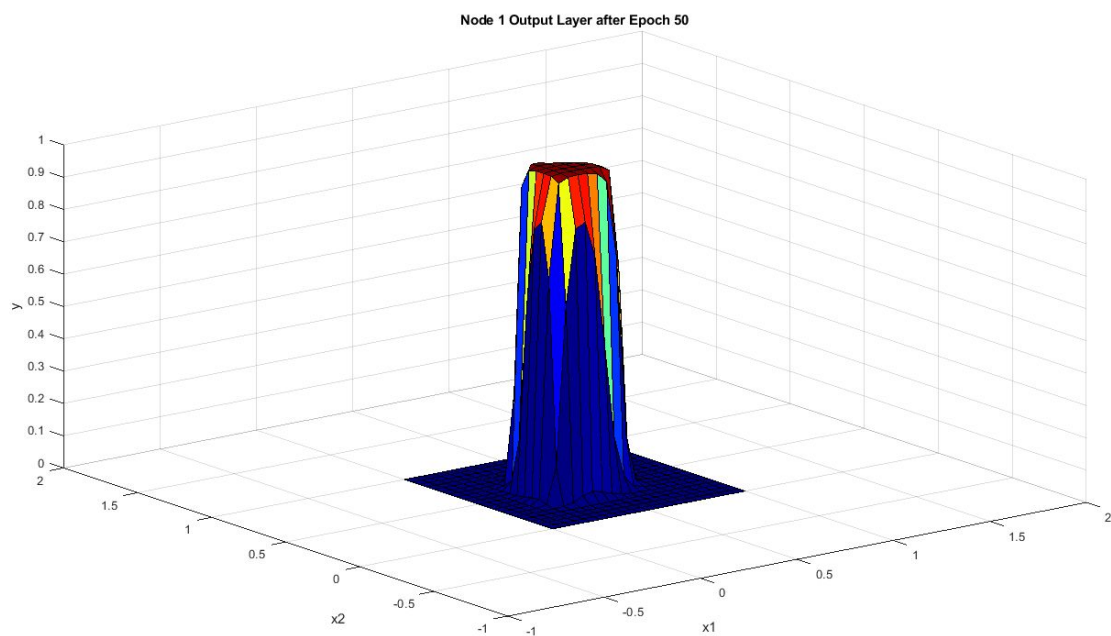
Hidden layer 1



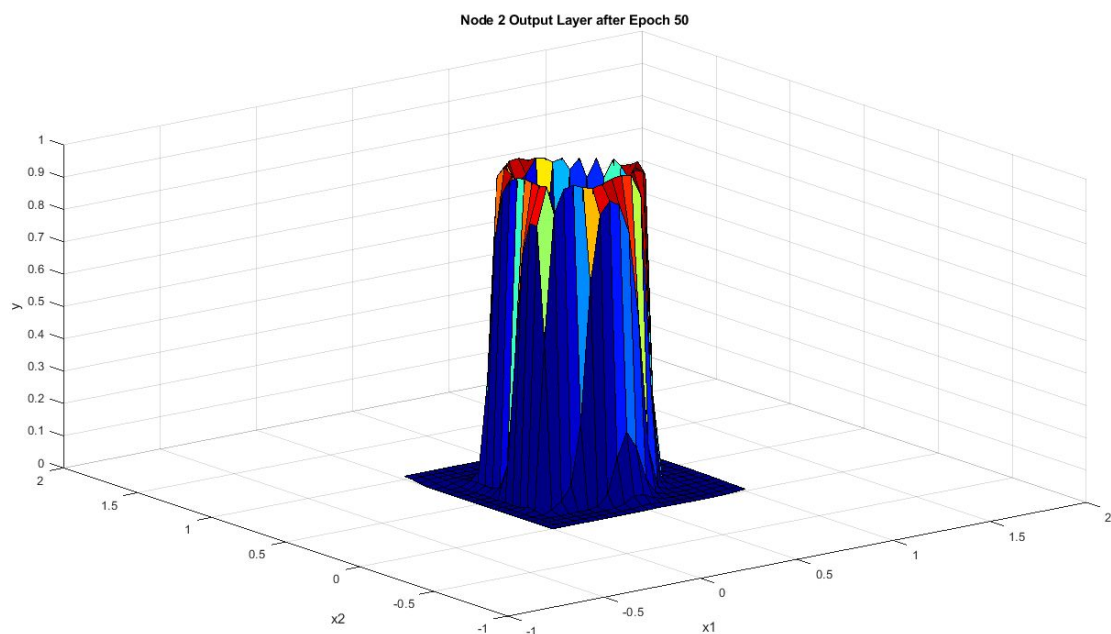
Hidden layer 2



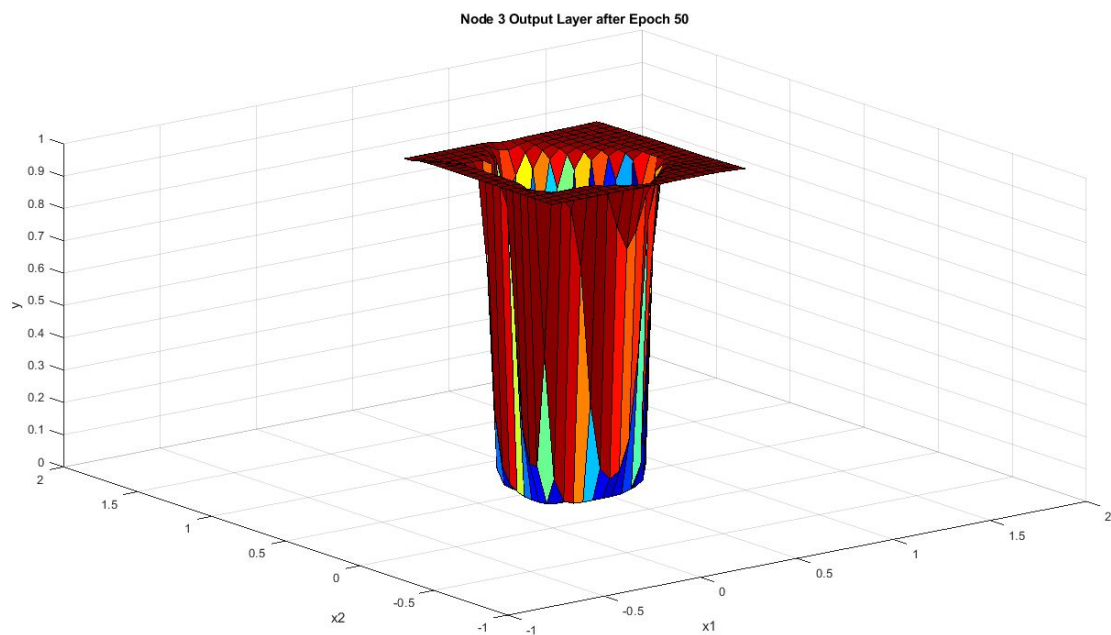
Node 1 of output layer



Node 2 of output layer

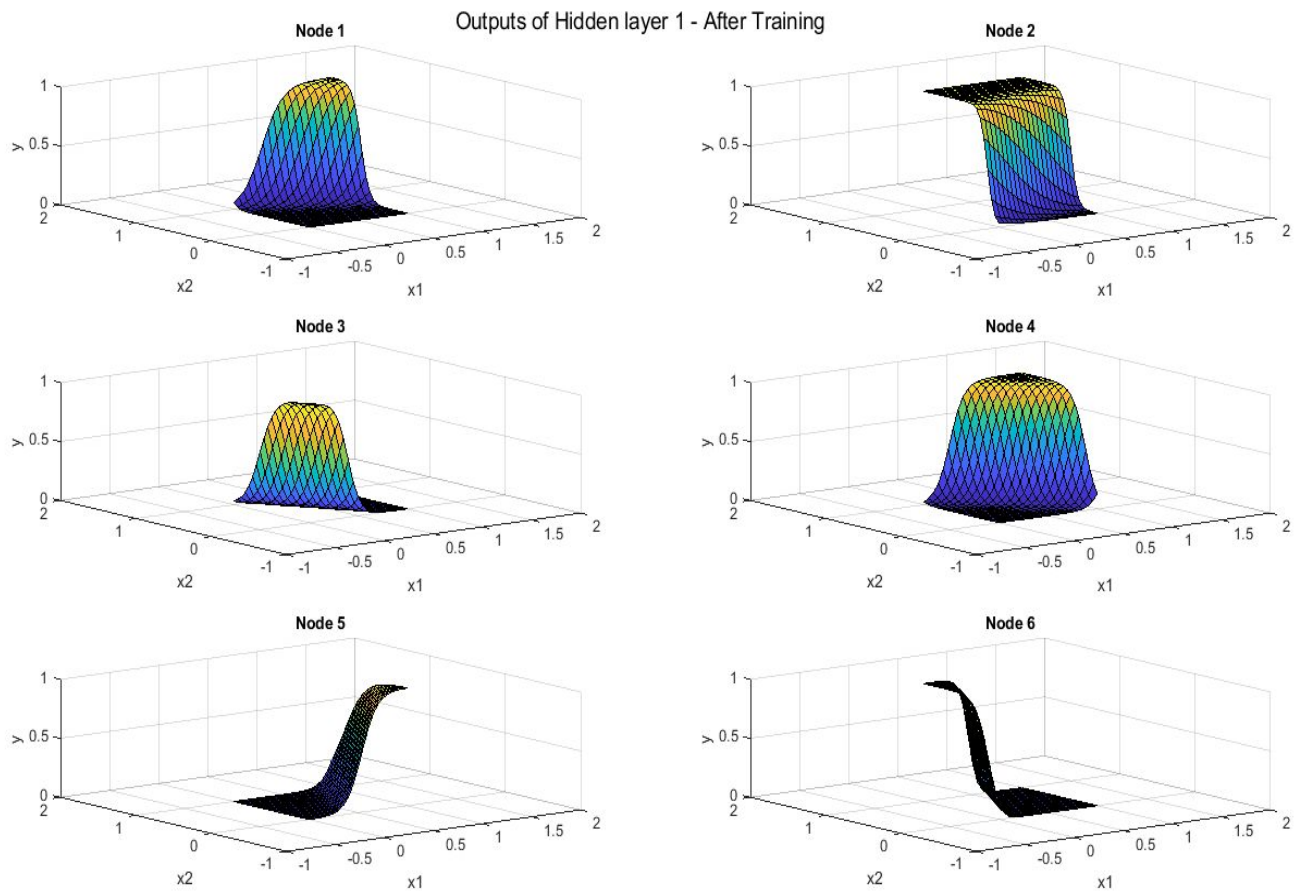


Node 3 of output layer

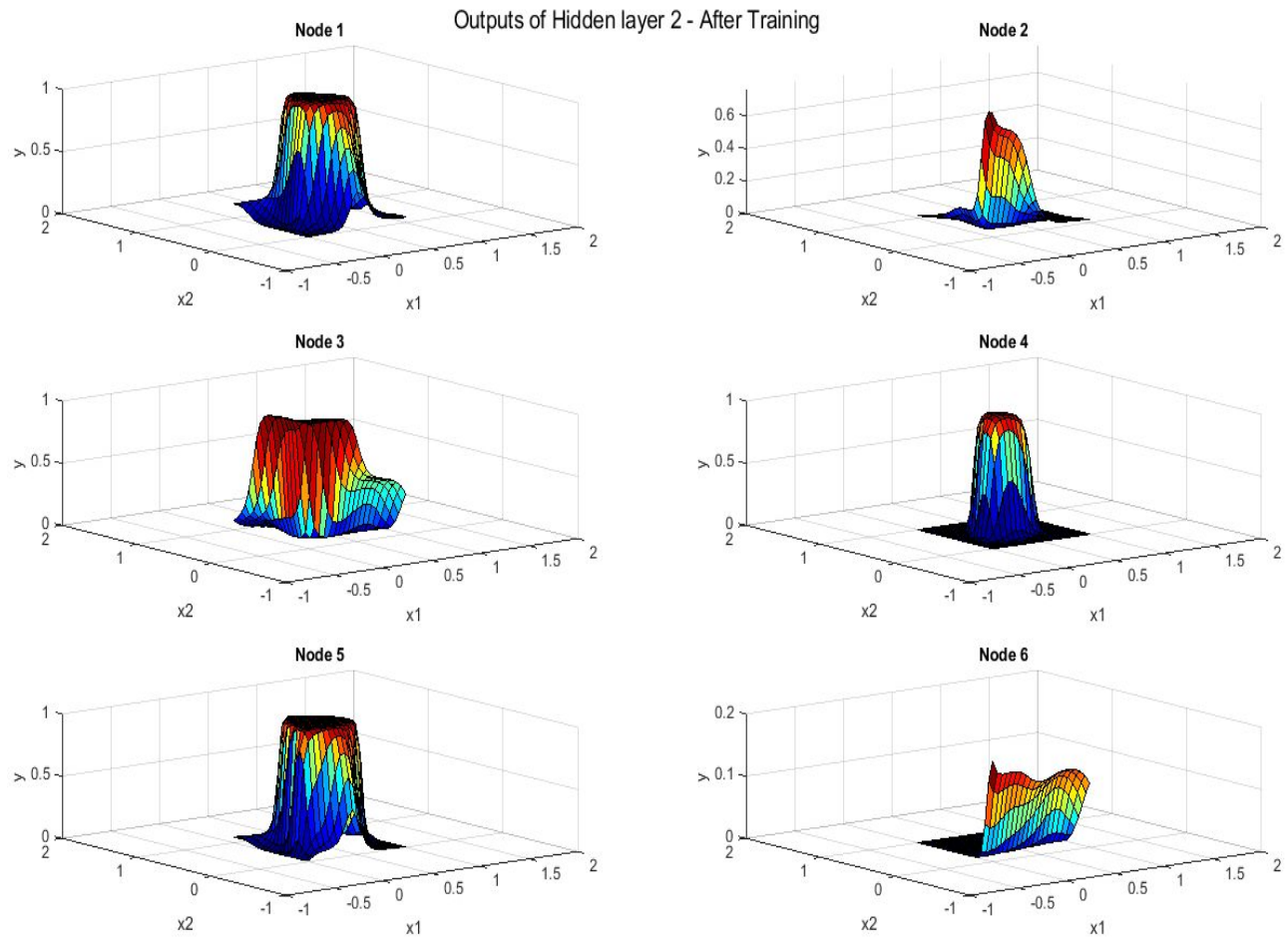


(9) Surfaces of outputs of nodes in hidden layers and output layer after training is stopped

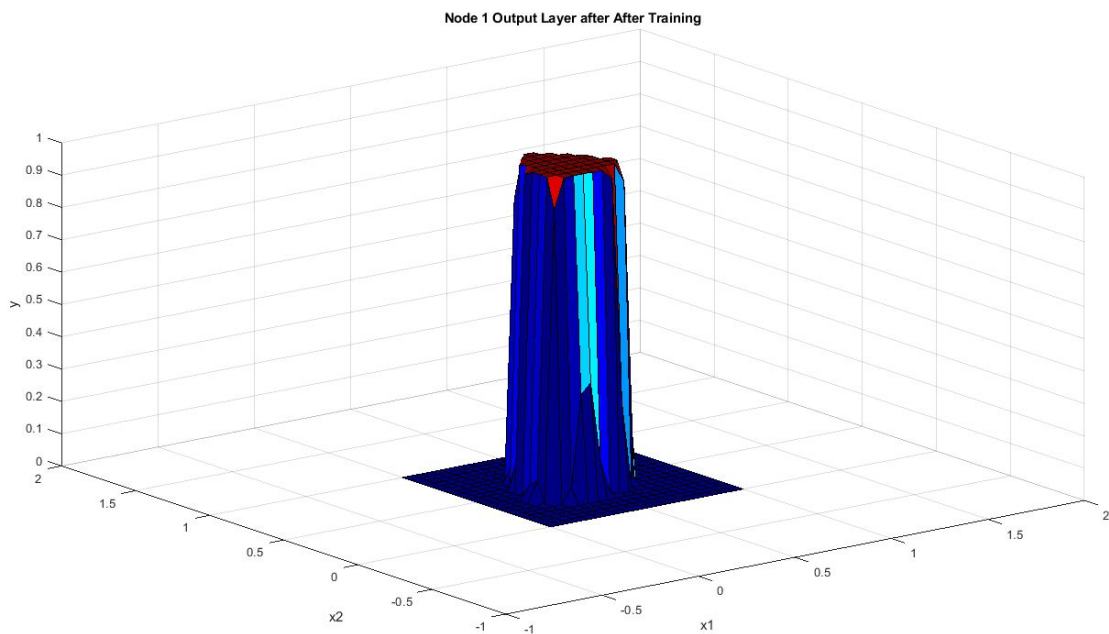
Hidden layer 1



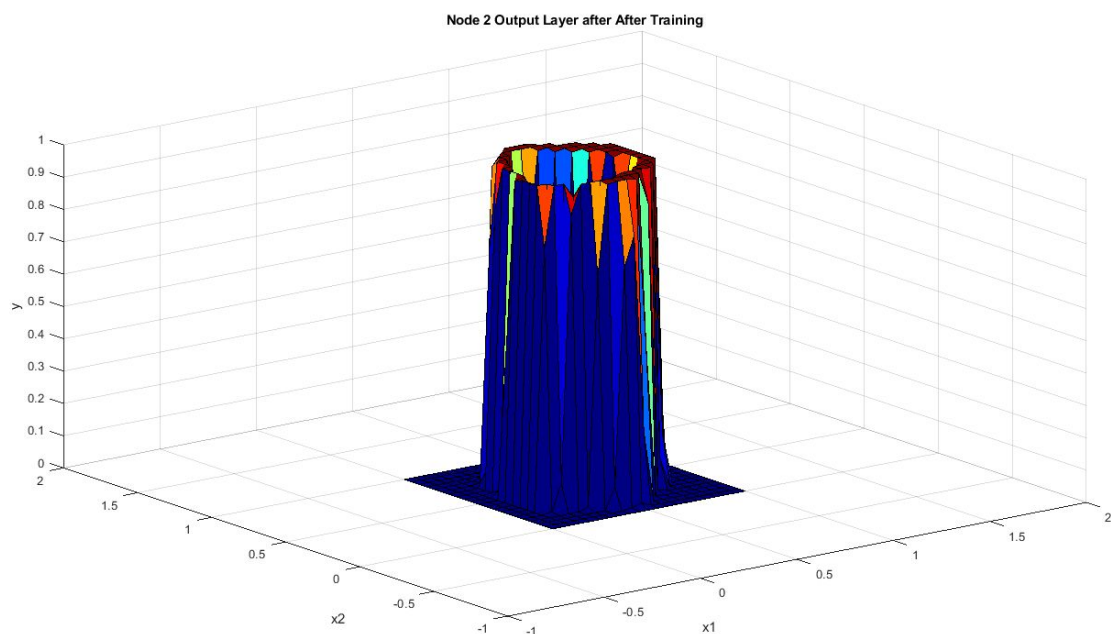
Hidden layer 2



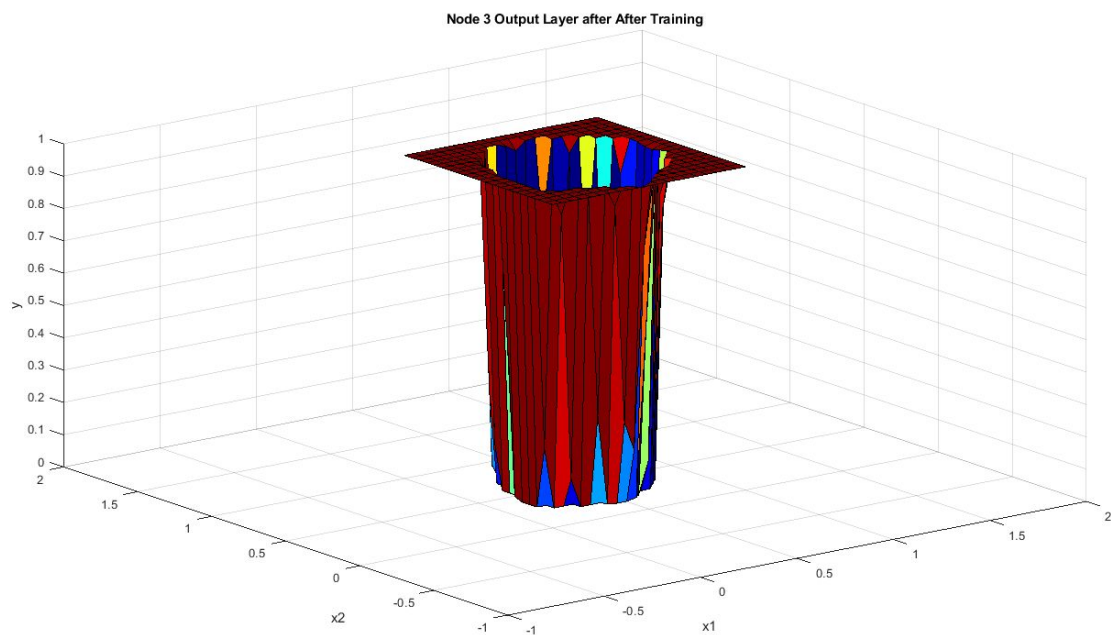
Node 1 of output layer



Node 2 of output layer



Node 3 of output layer



3. Image Classification

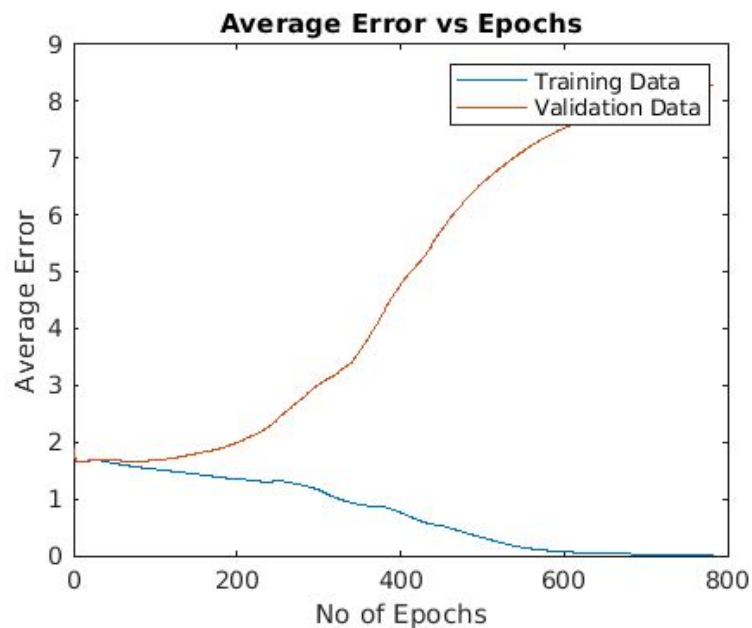
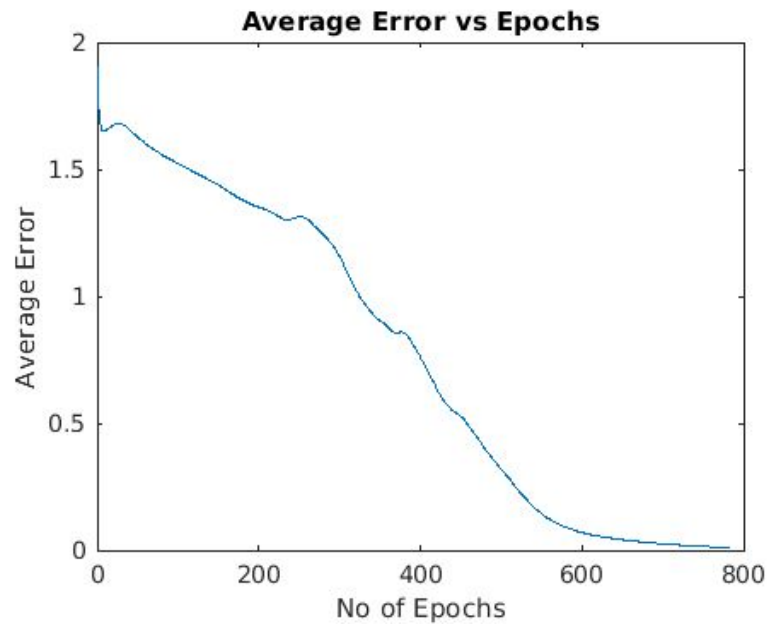
Algorithm: Principal Component Analysis (PCA) was used to reduce the dimension of extracted features from 512 to 30.

Stopping criteria: $\text{abs}(\text{average error difference between subsequent epochs}) < 10^{-4}$) and average error < 0.11

(1) Delta rule

$$\eta = 0.08;$$

- a. Plot of average error on training data vs Epoch



b. Confusion matrix for training data and validation data

Training Data

222	0	0	0	0
0	204	0	0	0
0	0	212	0	0
0	0	0	208	0
0	0	0	0	204

Accuracy = 100 %

Validation Data

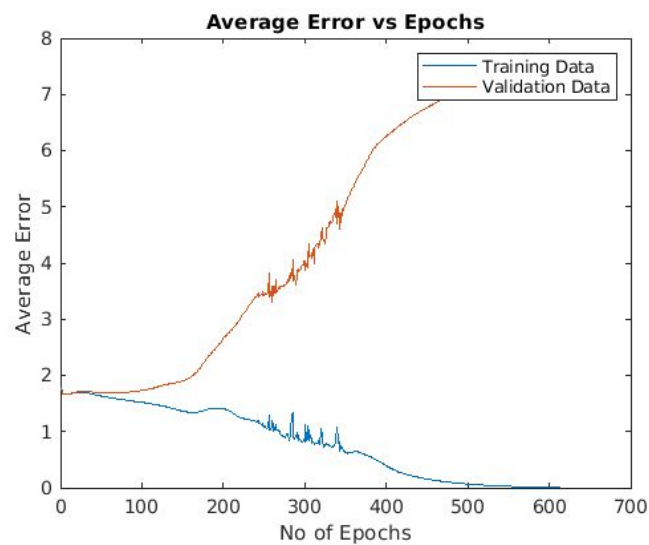
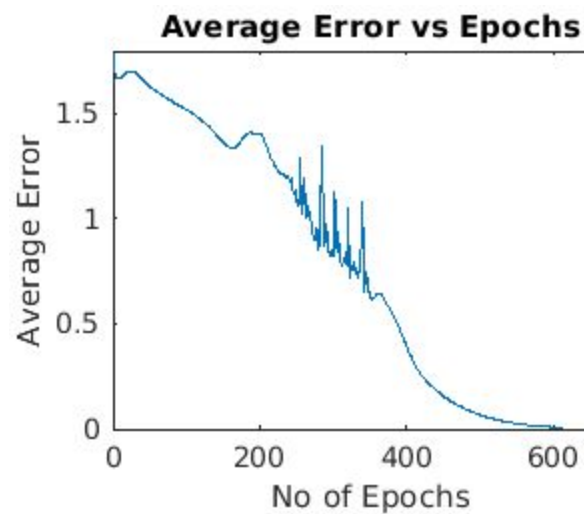
17	14	17	13	17
11	17	21	16	31
22	13	19	18	16
14	15	25	18	20
16	27	11	18	24

Accuracy = 21.11%

(2) Generalized Delta rule

$$\eta = 0.08, \quad \alpha = 0.5$$

a. Plot of average error on Training data vs Epoch



b. Confusion matrix for training data and validation data

Training Data

222	0	0	0	0
0	204	0	0	0
0	0	212	0	0
0	0	0	208	0
0	0	0	0	204

Accuracy = 100 %

Validation Data

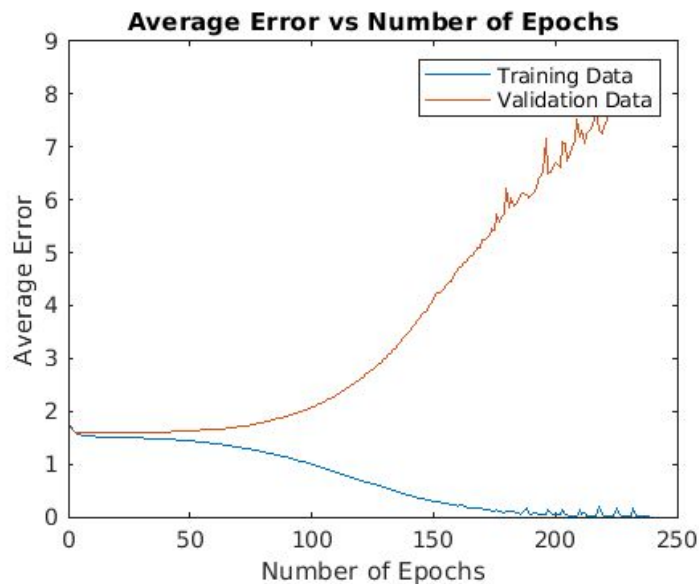
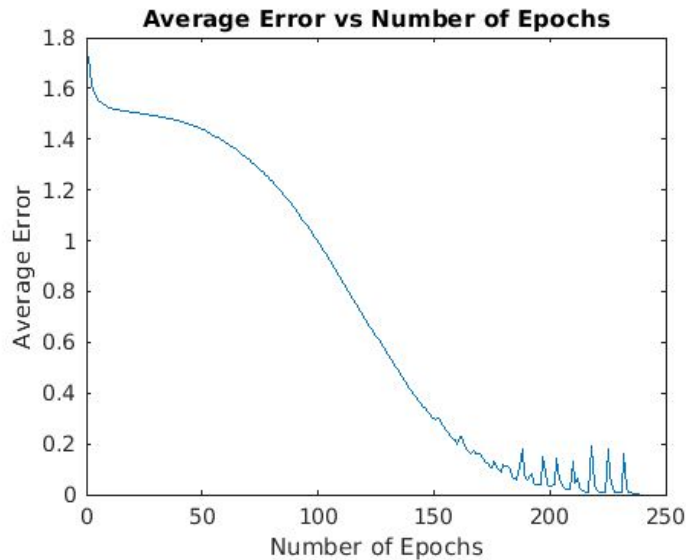
15	12	20	12	19
21	32	12	17	14
16	13	23	19	17
17	18	19	21	17
20	17	19	15	25

Accuracy = 25.78%

(3) Adam Method

$$\eta = 0.01, \quad \epsilon = 10^{-8}, \quad \rho_1 = 0.9, \quad \rho_2 = 0.999$$

a. Plot of average error on training data vs Epoch



b. Confusion matrix for training data and test data

Training Data

222	0	0	0	0
0	204	0	0	0
0	0	212	0	0
0	0	0	208	0
0	0	0	0	204

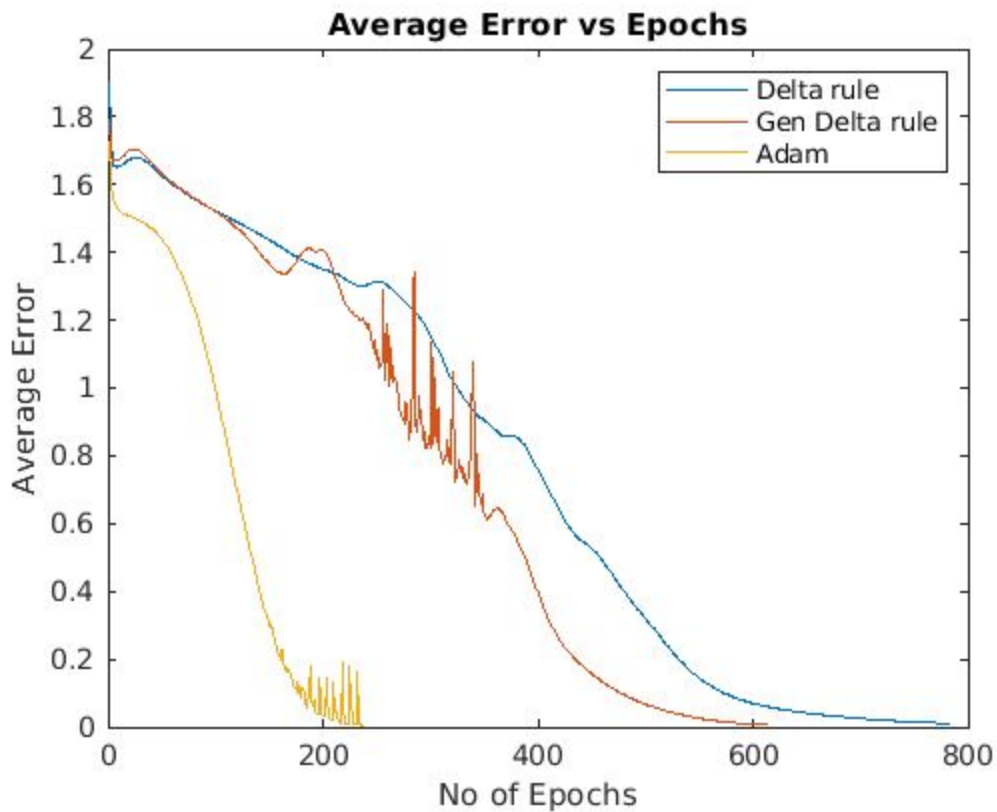
Accuracy = 100 %

Validation Data

9	15	22	5	27
14	20	13	10	39
13	11	25	13	26
15	21	26	10	20
13	14	22	8	39

Accuracy = 22.89%

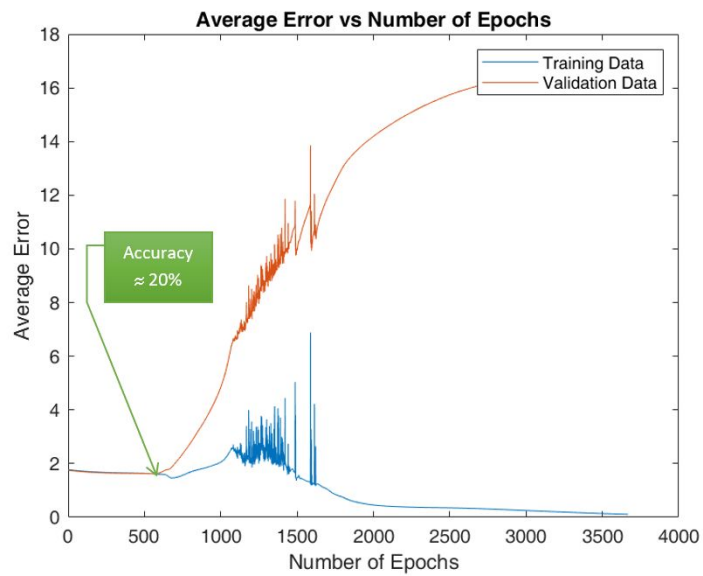
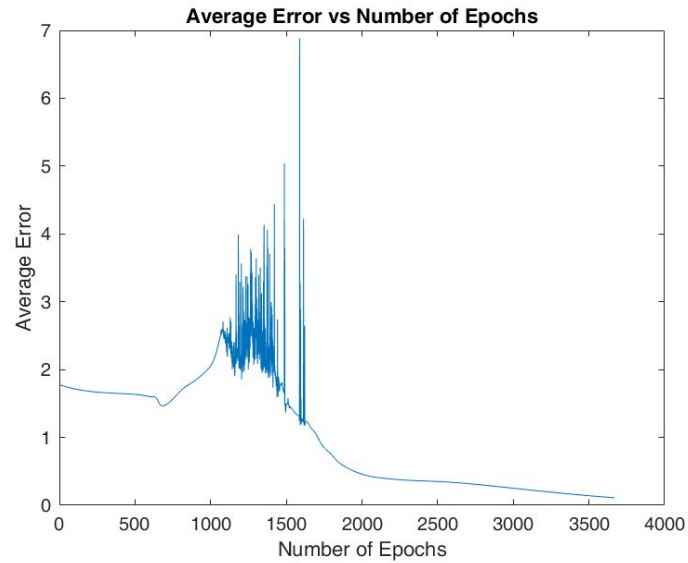
(4) Comparison of all the three learning methods



The Adam method gives the fastest and smoothest convergence compared to other methods. Also, it can be observed that the generalized delta rule is faster than the delta rule, but it does not result in a smooth convergence.

In the above model, it can be observed that the error for validation data is increasing for every epoch. This could be due to less number of training examples. So, the model is overfitting the training data. Hence, various configurations of the same model with different numbers of features and different numbers of nodes in hidden layers were tried. But validation accuracy didn't improve. Consequently, other models were tried out and the results of the same are presented below.

1. All the extracted 512 features were used to train the model without reducing the dimension.
 - a. Plot of average error on training data vs Epoch



b. Confusion Matrix

Training Data

216	0	5	1	0
0	204	0	0	0
0	0	212	0	0
0	0	0	208	0
2	4	13	8	177

Accuracy = 96.86 %

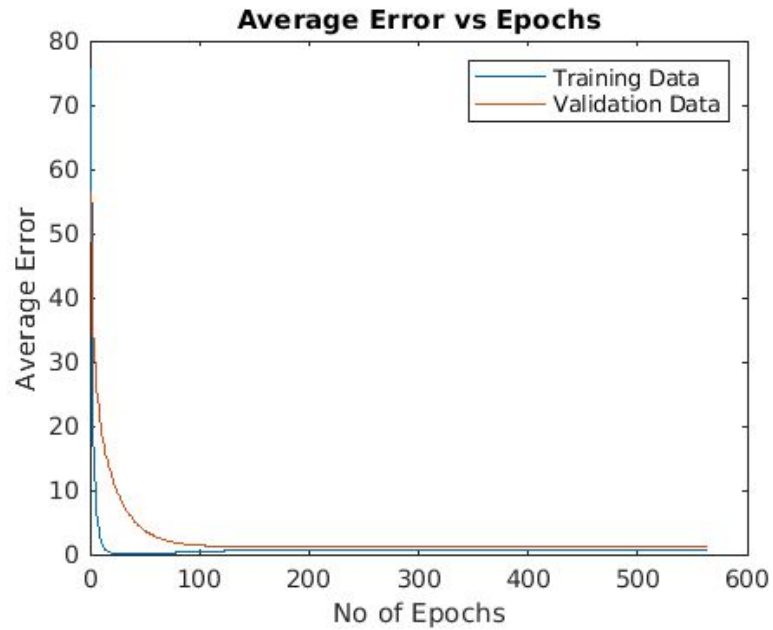
Validation Data

17	14	16	16	15
23	29	12	14	18
9	16	26	28	9
14	13	31	21	13
18	12	19	25	22

Accuracy = 25.56%

2. Image data was directly used to train the model. PCA was carried out to reduce the dimension of the input features from 3072 to 360.

a. Plot of average error on training data vs Epoch



b. Confusion Matrix

Training Data

166	5	6	14	16
4	203	7	1	6
6	26	146	28	9
9	6	22	152	13
19	14	4	16	152

Accuracy = 78 %

Validation Data

49	13	4	10	17
8	59	0	2	10
7	14	43	12	9
7	5	24	53	9
16	10	3	19	47

Accuracy = 55.78%