Comparison of Training and Activation Functions In Single and Multi-layer Neural Network

Aytekin YILDIZHAN
aytekinyildizhan@hacettepe.edu.tr
N18147923
CMP684 Neural Networks Project Report

***Abstract*—This project has two parts. First part compares six BP training functions in single layer feedforward network and compares four activation functions in multi-layer feedforward network. Second part compares six BP training functions in single layer pattern recognition network and compares four activation functions in multi-layer pattern recognition network. In single layer feedforward network, `trainlm` shows a better performance whereas `traingd` and `traingdm` show a worse performance. `trainlm` is used for multi-layer feedforward network. In multi-layer feedforward network, `logsig` shows a better performance whereas `purelin` shows a worse performance. In single layer pattern recognition network, `trainlm` shows better performance while `traingd` shows a worse performance. `trainlm` is used for multi-layer pattern recognition network. In multi-layer pattern recognition network, `tansig` shows a better performance whereas `hardlim` shows a worse performance.**

Keywords—feedforward network; pattern recognition network, back-propagation algorithm, activation function

I. INTRODUCTION

This study consists of two parts. First part is that analyzing single and multi-layer feedforward network and the other part is that analyzing single and multi-layer pattern recognition network. We use two data sets form Neural Network Datasets example in MATLAB. First one is house dataset for feedforward network. Second data set is cancer dataset for pattern recognition network. For these neural networks training, the first 60% of data is used to analyze the six BP training functions, whereas the next 20% is used for the validation step and the remaining 20% for the testing step.

House dataset includes 13 attributes of 506 neighborhoods so the feedforward network has 13 inputs. The output is the median values of owner-occupied homes so this network has 1 output.

There are several type of BP training functions. In this work, six BP training functions are applied which are Levenberg-Marquadt Back-Propagation (`trainlm`), Scaled Conjugated Gradient Back-Propagation (`trainscg`), Gradient Descent with Momentum (`traingdm`), Gradient Descent with Adaptive Learning Rate Back-Propagation (`traingda`), Gradient Descent with Momentum and Adaptive Learning Rate Back-Propagation (`traingdx`), and Gradient Descent Back-Propagation algorithm (`traingd`). The errors are measured by Mean Squared Error (MSE).

For the single layer feedforward network, four activation functions are applied. These functions are hyperbolic tangent sigmoid transfer function (`tansig`), log-sigmoid transfer function (`logsig`), linear transfer function (`purelin`), and hard-limit transfer function (`hardlim`). Output neuron is a linear function. BP training function with the least MSE value will be selected in single layer feedforward network and used in multi-layer feedforward neural network.

In multi-layer feedforward network consist of two hidden layers and the activation functions in hidden layers are the same. Activation function with the least MSE value will be determine.

Cancer dataset includes 9 attributes of 699 biopsies so the pattern recognition network has 9 inputs. The output is the cancer type, which are benign or malignant. Therefore this network has 2 outputs.

BP training functions are the same as feedforward network. BP training function with the least MSE value will be selected in single layer pattern recognition network and used in multi-layer pattern recognition network.

For the single layer pattern recognition network, four activation functions are applied. These functions are hyperbolic tangent sigmoid transfer function (`tansig`), log-sigmoid transfer function (`logsig`), linear transfer function (`purelin`), and hard-limit transfer function (`hardlim`). Output neuron is a soft max transfer function (`softmax`).

In multi-layer feedforward network consist of two hidden layers and the activation functions in hidden layers are the same. Activation function with the least MSE value will be determine.

The rest of this project is organized as follows. Section 2 and section 3 single layer and multi-layer feedforward networks are analyzed, respectively. Section 4 and section 5 single layer and multi-layer pattern recognition networks are analyzed, respectively. Finally, the conclusions and future works are given in Section 6.

II. SINGLE LAYER FEEDFORWARD NETWORK

Firstly of all, during the performance of this project, we choose the number of neurons (R) of the hidden layer. We

consider the cases with R = 2, 4, 8, 16, 32, 64 and 128 neurons. The default values are initialized by MATLAB code `'feedforwardnet'`.

Secondly, every individual training trial is terminated after the completion of 5000 epochs, the epoch error increases for 10 consecutive epochs, MSE value is equal to zero or minimum gradient value is reached. We follow a grading policy such that the activation function resulting in the least value is assigned a grade 5 and that associated to the highest value is assigned a grade 0. Tables show that final MSE value and their grade. For each one of the different hidden layer configurations, we obtain the Tables 1–7. Also the least MSE value colored red and the highest MSE value colored blue.

TABLE 1. THE MSE RESULTS OF R = 2 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	5 20.3813	5 21.3617	4 45.1061	4 94.0354
trainscg	4 21.1756	4 21.8581	5 30.9345	3 98.5158
traingdm	0 9.6024+e03	1 98.8814	1 517.3559	0 153.6213
traingda	3 72.0005	2 91.8292	2 82.7432	5 87.5449
traingdx	2 92.7261	3 78.2495	3 64.3427	2 100.1529
traingd	1 183.8205	0 510.0609	0 924.7337	1 137.4428

TABLE 2. THE MSE RESULTS OF R = 4 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	5 22.0706	5 27.1154	2 135.3580	5 71.1083
trainscg	4 22.1934	4 31.1715	5 40.3383	4 81.6739
traingdm	0 561.2484	0 121.5117	1 2.3324e+03	0 701.2672
traingda	3 22.8789	3 36.5782	4 58.2074	3 84.9443
traingdx	2 96.4939	2 47.6162	3 106.0609	2 131.6657
traingd	1 163.2470	1 100.5294	0 4.7553e+03	1 353.3106

TABLE 3. THE MSE RESULTS OF R = 8 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	5 22.5964	5 25.2127	2 197.4523	4 63.6182
trainscg	2 47.6721	3 79.3886	5 29.9353	3 63.6224
traingdm	0 887.4148	1 443.7812	1 2.9355e+03	0 555.0189
traingda	3 35.2161	2 96.0565	4 50.5489	2 96.3413
traingdx	4 28.8046	4 43.2653	3 80.4189	5 61.2486
traingd	1 760.4936	0 741.1032	0 8.7055e+03	1 468.6074

TABLE 4. THE MSE RESULTS OF R = 16 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	4 38.8315	5 25.6082	3 153.5230	5 57.0224
trainscg	3 52.3016	4 39.7159	5 32.8374	3 81.5470
traingdm	1 1.1280e+03	0 1.3310e+03	1 1.6114e+03	0 1.3377e+03
traingda	5 33.8854	3 67.4059	4 54.3428	4 61.0202
traingdx	2 101.4170	2 167.1947	2 299.4116	2 335.9887
traingd	0 1.9324e+03	1 1.2478e+03	0 1.0055e+04	1 657.1056

TABLE 5. THE MSE RESULTS OF R = 32 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	5 29.9849	5 34.3697	2 153.5230	5 52.4716
trainscg	4 39.5392	4 40.4445	5 54.7467	4 55.5369
traingdm	1 578.7976	0 1.5472e+03	1 1.4482e+04	1 339.6279
traingda	3 39.8484	3 58.9965	4 58.4430	2 88.0212
traingdx	2 103.5374	2 100.0130	3 129.7442	3 83.1896

Traingd	0 770.1286	1 513.3648	0 1.7949e+04	0 1.7335e+03

TABLE 6. THE MSE RESULTS OF R = 64 NEURONS

	tansig	**logsig**	**purelin**	**hardlim**
trainlm	5 80.0858	3 85.2589	2 175.7452	3 84.8835
trainscg	4 96.0924	5 47.7481	4 48.9041	4 46.9110
traingdm	0 1.3424e+03	1 646.3897	1 1.1226e+04	1 2.3383e+04
traingda	3 128.5532	4 81.2221	3 87.0048	2 481.9711
traingdx	2 668.7208	2 431.0286	5 39.6045	5 32.8017
traingd	1 1.316e+03	0 1.0575e+04	0 1.4363e+04	0 6.9085+04

TABLE 7. THE MSE RESULTS OF R = 128 NEURONS

	tansig	**logsig**	**purelin**	**hardlim**
trainlm	5 257.2756	5 230.5944	4 143.0015	3 480.2362
trainscg	4 393.9713	4 452.9744	5 72.5260	5 79.1465
traingdm	0 6.8431e+03	0 4.0262e+03	1 3.3459e+04	0 2.6194e+04
traingda	3 427.9167	2 1.6162e+03	2 1.031e+04	2 2.2613+03
traingdx	2 802.3621	3 653.2820	3 191.5571	4 452.7661
traingd	1 5.6249e+03	1 2.7264e+03	0 5.0728e+04	1 5.9739e+03

After these calculations, Table 8 shows the comparison of total grade of each case is presented.

TABLE 8. OVERALL COMPARİSON OF THE RESULTS

R=	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**Average**
trainlm	18	17	16	17	17	13	17	16.42
trainscg	16	17	13	15	17	17	18	16.14
traingdm	2	1	2	2	3	3	1	2
traingda	12	13	11	16	12	12	9	12
traingdx	10	9	16	8	10	14	12	11.28
traingd	2	3	2	2	1	1	3	2

As can be seen from Table 8, the highest grade of training function is `trainlm` whereas the least grade of training function is both `traingd` and `traingdm`. As a result, `trainlm` is selected for multi-layer feedforward network.

III. MULTI-LAYER FEEDFORWARD NETWORK

In this section, multi-layer feedforward network is applied to house dataset and `trainlm` function is used. There are two hidden layers and activation functions in hidden layers are the same. In Table 9-15, MSE values are graded from least (3) to highest (0). Least MSE value colored red and highest MSE value colored blue.

TABLE 9. THE MSE RESULTS OF R = 2X2 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	2 36.8553	3 30.7514	1 57.7732	0 94.6545

TABLE 10. THE MSE RESULTS OF R = 4X4 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	3 29.6236	2 30.6606	0 381.7470	1 65.907

TABLE 11. THE MSE RESULTS OF R = 8X8 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	3 33.7226	2 44.62	0 152.2528	1 136.6329

TABLE 12. THE MSE RESULTS OF R = 16X16 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	2 57.0822	3 22.1930	0 128.8109	1 73.4676

TABLE 13. THE MSE RESULTS OF R = 32X32 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	3 59.6204	1 77.9766	0 149.7821	2 68.0214

TABLE 14. THE MSE RESULTS OF R = 64X64 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	0 282.4904	2 92.0429	1 153.5230	3 77.5472

TABLE 15. THE MSE RESULTS OF R = 128X128 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	0 3.9791e+03	1 723.1095	3 153.4826	2 297.7083

After these calculations, Table 16 shows the comparison of total grade of each case is presented.

TABLE 16. OVERALL COMPARİSON OF THE RESULTS

	tansig	logsig	purelin	hardlim
Average	1.85	2	0.71	1.42

As you can see from Table 16, `logsig` function is the highest grade whereas `purelin` is the least grade. Therefore for house dataset, `logsig` function is the best activation function in the multi-layer feedforward neural network using two hidden layers.

IV. SINGLE LAYER PATTERN RECOGNITION NETWORK

This second experiment criteria is same as the single layer feedforward network. For each one of the different hidden layer configurations, we obtain the Tables 17–23.

TABLE 17. THE MSE RESULTS OF R = 2 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	3 0.1421	4 0.1409	4 0.1389	5 0.1436
trainscg	4 0.1410	5 0.1387	5 0.1383	1 0.2759
traingdm	1 0.1535	1 0.2736	1 0.1446	0 0.2811
traingda	3 0.1421	2 0.1441	3 0.1398	3 0.2423
traingdx	5 0.1382	3 0.1422	3 0.1398	4 0.1450
traingd	0 0.1881	0 0.2859	0 0.1470	2 0.2352

TABLE 18. THE MSE RESULTS OF R = 4 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	5 0.1406	5 0.1374	3 0.1405	5 0.1532
trainscg	2 0.1449	4 0.1382	2 0.1433	4 0.1879
traingdm	1 0.1529	1 0.2007	1 0.145	2 0.214
traingda	4 0.1428	2 0.1399	5 0.1391	0 0.2447
traingdx	3 0.1434	3 0.1384	4 0.1397	1 0.2244
traingd	0 0.1531	0 0.2563	0 0.1460	3 0.203

TABLE 19. THE MSE RESULTS OF R = 8 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	0 0.203	5 0.1374	3 0.1389	3 0.1507
trainscg	3 0.144	4 0.1378	2 0.1405	4 0.1491
traingdm	2 0.1493	1 0.1681	0 0.1497	1 0.176
traingda	5 0.1374	2 0.1387	4 0.1388	2 0.1647
traingdx	4 0.1421	3 0.138	5 0.1386	5 0.1399
traingd	1 0.1515	0 0.1701	1 0.1461	0 0.2152

TABLE 20. THE MSE RESULTS OF R = 16 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	2 0.1432	5 0.1391	0 0.1459	4 0.1442
trainscg	3 0.1381	3 0.1395	5 0.1384	5 0.1365
traingdm	0 0.1467	0 0.4543	3 0.1431	1 0.1638
traingda	4 0.1375	4 0.1392	1 0.1438	3 0.153
traingdx	5 0.1374	3 0.1395	4 0.1404	2 0.1618
traingd	1 0.1437	1 0.1701	3 0.1431	0 0.1691

TABLE 21. THE MSE RESULTS OF R = 32 NEURONS

	tansig	logsig	purelin	hardlim
trainlm	5 0.136	2 0.1415	5 0.1386	3 0.1410
trainscg	4 0.1361	3 0.1401	3 0.1419	5 0.1399
traingdm	1 0.1434	1 0.1521	1 0.1436	1 0.153
traingda	3 0.1379	5 0.138	4 0.1405	4 0.1409
traingdx	2 0.1395	4 0.1389	2 0.1422	2 0.1421

traingd	0 0.1438	0 0.1542	1 0.1436	0 0.4505

TABLE 22. THE MSE RESULTS OF R = 64 NEURONS

	tansig	**logsig**	**purelin**	**hardlim**
trainlm	5 0.1341	5 0.1319	4 0.1417	5 0.137
trainscg	4 0.1359	4 0.1387	5 0.1384	4 0.1396
traingdm	1 0.1433	1 0.1645	2 0.1425	0 0.15
traingda	2 0.1408	3 0.14	1 0.1426	3 0.1414
traingdx	3 0.14	0 0.2172	0 0.153	2 0.1458
traingd	0 0.1498	2 0.146	4 0.1417	1 0.1479

TABLE 23. THE MSE RESULTS OF R = 128 NEURONS

	tansig	**logsig**	**purelin**	**hardlim**
trainlm	5 0.1348	5 0.1354	0 0.1598	5 0.1341
trainscg	3 0.138	4 0.1427	1 0.1504	2 0.1426
traingdm	0 0.1425	3 0.1438	5 0.1421	1 0.1444
traingda	4 0.1376	2 0.1445	2 0.1444	4 0.1363
traingdx	2 0.1421	1 0.4526	4 0.1428	3 0.1365
traingd	2 0.1421	0 0.4528	3 0.1443	0 0.15

After these calculations, Table 24 shows the comparison of total grade of each case is presented.

TABLE 24. OVERALL COMPARISON OF THE RESULTS

R=	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**Average**
trainlm	16	18	11	11	15	19	15	15
trainscg	15	12	13	16	15	17	10	14
traingdm	3	5	4	4	4	4	9	4.14
traingda	11	11	13	12	16	9	12	12
traingdx	15	11	17	14	10	5	10	11.71
traingd	2	3	2	5	1	7	5	3.57

As can be seen from Table 24, the highest grade of training function is `trainlm` whereas the least grade of training function is `traingd`. As a result, `trainlm` is selected for multi-layer pattern recognition network.

V. MULTI-LAYER PATTERN RECOGNITION NETWORK

In this section, multi-layer pattern recognition network is applied to cancer dataset and `trainlm` function is used. There are two hidden layers and activation functions in hidden layers are the same. In Table 25-31, MSE values are graded from least (3) to highest (0). Least MSE value colored red and highest MSE value colored blue.

TABLE 25. THE MSE RESULTS OF R = 2X2 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	1 0.1468	3 0.1375	2 0.1407	0 0.1970

TABLE 26. THE MSE RESULTS OF R = 4X4 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	2 0.1398	1 0.1438	3 0.1393	0 0.2724

TABLE 27. THE MSE RESULTS OF R = 8X8 NEURONS

	tansig(x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	1 0.1447	2 0.1435	3 0.1401	0 0.1519

TABLE 28. THE MSE RESULTS OF R = 16X16 NEURONS

	tansig (x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	3 0.1339	2 0.1423	1 0.1434	0 0.1534

TABLE 29. THE MSE RESULTS OF R = 32X32 NEURONS

	tansig (x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	3 0.1368	2 0.1413	0 0.1425	1 0.1421

TABLE 30. THE MSE RESULTS OF R = 64X64 NEURONS

	tansig (x2)	**logsig(x2)**	**purelin(x2)**	**hardlim(x2)**
trainlm	3 0.137	2 0.1371	0 0.1447	1 0.138

TABLE 31. THE MSE RESULTS OF R = 128X128 NEURONS

	tansig(x2)	logsig(x2)	purelin(x2)	hardlim(x2)
trainlm	3 0.1338	1 0.1393	0 0.1464	2 0.1359

After these calculations, Table 32 shows the comparison of total grade of each case is presented.

TABLE 32. OVERALL COMPARİSON OF THE RESULTS

	tansig	logsig	purelin	hardlim
Average	2.28	1.85	1.28	0.57

As you can see from Table 32, `tansig` function is the highest grade whereas `hardlim` is the least grade. Therefore for cancer dataset, `tansig` function is the best activation function in the multi-layer pattern recognition neural network using two hidden layers.

VI. CONCLUSION

This project presents a comparison of BP training and neuronal activation functions. In first experiment, we consider four different neuronal activation functions, six BP training functions and having 2, 4, 8, 16, 32, 64, 128 hidden neurons for a single layer feedforward network. `trainlm` displays a better performance among six BP training functions. By using `trainlm`, we consider four different neuronal activation functions and two hidden layers with 2x2, 4x4, 8x8, 16x16, 32x32, 64x64, 128x128 neurons for a multi-layer feedforward network. `logsig` displays a better performance whereas `purelin` displays a worse performance among four neural activation functions.

In second experiment, we consider four different neuronal activation functions, six BP training functions and having 2, 4, 8, 16, 32, 64, 128 hidden neurons for a single layer pattern recognition network. `trainlm` displays a better performance among six BP training functions. By using `trainlm`, we consider four different neuronal activation functions and two hidden layers with 2x2, 4x4, 8x8, 16x16, 32x32, 64x64, 128x128 neurons for a multi-layer pattern recognition network. `tansig` displays a better performance whereas `hardlim` displays a worse performance among four neural activation functions.

For small sized single layer feedforward networks, `trainlm` seem to be the best approach while `trainscg` seems to be the best approach for big sized single feedforward network.

For small sized single layer pattern recognition networks, `trainscg` seem to be the best approach while `trainlm` seems to be the best approach for big sized single layer pattern recognition network. `traingd` is the worst approach for both network configuration since it is too slow and its MSE value is too big.

For 128x128 hidden neurons in multi-layer pattern recognition networks, it lasted six hours. Thus, while the number of neurons or hidden layers increases, the elapsed time also increases. This approach may not be practical.

Future works may include different dataset, may increase the number of activation functions, the number of different training functions, hidden layers or neurons. Also different network configurations may be applied and compared to each other.