

CS 6375

ASSIGNMENT HW2

Names of students in your group:

Savan Amitbhai Visalpara (sxv180069)

Number of free late days used: 0

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

<https://scikit-learn.org/stable/documentation.html>

<https://stackoverflow.com>

Note: Please find **part1-hw2.pdf** for part-1 answers.

Content:

1. Introduction
2. Effect of Activation Functions
3. Experiments with Different Parameters

Introduction

This assignment asks to implement neural network with backpropagation for three types of activation functions. This brief report summarizes the results.

The given neural network is a three-layer neural network with 2 hidden layers. The implementation of sigmoid is provided for reference and it is also used by default. The following two functions are also implemented as part of this assignment.

- Tanh
- Relu

The effect of different activation functions is explained in the following section.

Effect of Different Activation Functions

I started the experiments with the default activation function- sigmoid. After some parameter tweaking, I got the total error of 0.75 and test error of 0.95.

I also tried with different activation functions but I could not get better results than this. Ideally, RELU should give better results than sigmoid but it didn't in this case. I think the reason might be the error function or the design of the neural network itself.

Generally, RELU is used with classification problem where error function is cross-entropy or related function but in this case it mean squared error. Basically, we trained a regression network for a classification task.

```
[0.174249138 0.95]]
After 180000 iterations, the total error is 0.7576108650939424
The final weight vectors are (starting from input to output layers)
[[-1.53278816e-01  4.40709398e-01 -1.00150865e+00 -4.95343247e-01
 -7.18277135e-01 -5.46362254e-01  6.50886080e+00 -5.63467424e-03
 -2.03670765e-01 -1.56042590e-01]
 [-1.94294037e-01  3.71185817e-01 -6.17223142e-01  1.47942184e+00
 -1.10819543e+00  5.56384306e-02 -1.27990716e-01  8.80184436e-01
 -7.27436331e-01 -1.04518603e+00]
 [ 4.61445802e-01  9.36656343e-01 -3.76456866e-01  1.10691248e-01
  6.34479842e-01  2.67495514e+00 -9.93203506e-01 -3.17950852e+00
 -6.77189741e-01  2.71412009e-01]
 [-7.73614257e-01 -1.57505509e-01  8.97073791e-01 -5.83553477e-01
  2.65655677e-01  3.21743888e-01  2.56721341e-01 -2.75756674e-02
 -9.62878306e-01  2.65181702e-01]
 [ 1.00154118e+00  4.96380536e-01 -4.40846077e-01  5.00254475e-01
 -8.05105583e-01 -8.42996228e-02 -4.70668464e+00  1.17195534e+00
 -4.27238829e-01 -9.64101686e-01]
 [-9.46985439e-01  3.57876268e-01 -5.83557176e-01 -6.48083158e-01
  5.17570667e-02 -1.02033837e+00  8.98813537e-02 -2.03655612e-01
  8.41643549e-02  1.20819338e+00]]
[[-0.80609956 -0.17191913  0.39770095 -0.16775332 -0.91411479]
 [ 0.74833921  0.29352453 -1.24668808  0.22544016  0.64241673]
 [ 0.8043546  -0.72571171 -0.7183621  0.61661426 -0.20854938]
 [ 0.8844291  0.92155664 -1.72957108  0.01968966  2.15044126]
 [ 1.87441735  0.19229621  0.32994417 -0.9966078  0.88518309]
 [-1.24331173 -0.11574065  2.46495215  1.51270273 -2.45092438]
 [-3.02108243  0.55894057  2.51337589  0.90250275 -3.35797505]
 [-2.05865027  0.80248732  1.71201652 -0.37167046 -2.01942776]
 [ 1.80986042  0.08444418  0.16718409 -1.50584266  3.55393672]
 [-0.61151995 -0.97887447  1.36116381  0.08381974 -0.56027391]]
[[-4.77153433]
 [-0.05896698]
 [ 4.7338039 ]
 [ 2.51605564]
 [-6.48032274]]
Test Error: 0.9329151990004394
PS D:\UTD\Assignment\ML\HW2>
```

Experiments with Different Parameters

Following table shows the results with different parameters.

Iterations	Learning rate	Train error
180000	0.0004	0.75
190000	0.01	2.35
90000	0.01	3.29
50000	0.1	6.96
50000	0.001	1.63
50000	0.004	4.7
120000	0.0004	1.51