CS 6375
ASSIGNMENT 2

Names of students in your group:

1. Anshul Pardhi (ARP180012)
2. Ashwani Kashyap (AXK190033)

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.

- Lecture Slides: E-learning
- Dataset Source: *https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer/breast-cancer.data*

2.Programming Part Report

From the UCI ML repository, the breast cancer dataset was chosen to train and test the neural network. The dataset can be found here, while more information on the dataset can be found here. The dataset contains 9 input attributes and one output (class) attribute. It is a relatively small dataset with 286 instances.

The following parameters were tuned in order to attain better results:

- Learning Rate
- Max Iterations
- Training-Testing Data%

Training and testing error sum for different activation functions (sigmoid, tanh, ReLu) for different values of the tuning parameters mentioned above is listed in the tables below.

	Learning Rate: 0.05	
	Max Iterations: 1000	
	Training data %: 75%	
	Testing data %: 25%	
Activation Function	**Training Error Sum**	**Testing Error Sum**
Sigmoid	88.50282234130717	34.22431562572993
Tanh	178.90041754398618	71.88932346790031
ReLu	107.38700614574185	35.61299385425811

	Learning Rate: 0.5	
	Max Iterations: 1000	
	Training data %: 75%	
	Testing data %: 25%	
Activation Function	**Training Error Sum**	**Testing Error Sum**
Sigmoid	86.35884827755144	34.15727839185308
Tanh	267.25228846158893	87.75564880164411
ReLu	107.38700614574185	35.61299385425811

Learning Rate: 1.0
Max Iterations: 1000
Training data %: 80%
Testing data %: 20%

Activation Function	Training Error Sum	Testing Error Sum
Sigmoid	113.25986633436756	29.740055128498934
Tanh	303.1227633213832	70.09347667871174
ReLu	113.25993561603744	29.740064383962526

Learning Rate: 0.05
Max Iterations: 2000
Training data %: 90%
Testing data %: 10%

Activation Function	Training Error Sum	Testing Error Sum
Sigmoid	100.6611252214229	14.508402445464954
Tanh	181.04169816061201	46.52313159097612
ReLu	126.18814749780508	16.811852502194906

Learning Rate: 0.5
Max Iterations: 2000
Training data %: 80%
Testing data %: 20%

Activation Function	Training Error Sum	Testing Error Sum
Sigmoid	93.1720275989999	29.668119425618666
Tanh	239.97514451159736	70.56781061535165
ReLu	113.25993561603744	29.740064383962526

Learning Rate: 1.0
Max Iterations: 2000
Training data %: 90%
Testing data %: 10%

Activation Function	Training Error Sum	Testing Error Sum
Sigmoid	126.1881465311347	16.811852400157985
Tanh	214.3009079398663	24.571720225585032
ReLu	126.18814749780508	16.811852502194906

It is observed that both sigmoid as well as ReLu output the better training and testing error sum, (sigmoid being very very slightly better), followed by tanh. It is so because of the relatively smaller dataset (it contains only 286 instances). Plus, the network is not that deep (it contains only 2 hidden layers with 4 nodes in the first hidden layer and 2 nodes in the second hidden layer). Had the network been more deep, ReLu had been a better activation function.

It is observed that the best (though it might not be the best and we could obtain better results) training and testing error sum is obtained when the learning rate is 0.05, max iterations are 2000 and the train-test split is in the ratio 90%-10%.

Complete outputs for the above-mentioned tables are attached in the Outputs folder as separate files.