CS 6375
ASSIGNMENT _____3_____

Names of students in your group:

Nikhil Kumar Chandra (nxc161330)

Shruti Harihar (sxh164530)

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.

1) www.stackoverflow.com

2) Tom Mitchell Textbook

3) https://www.mimuw.edu.pl/~son/datamining/DM/4-preprocess.pdf

4) http://neuralnetworksanddeeplearning.com/chap2.html

Python Version: 2.7

Libraries used:

1. numpy
2. pandas

Assumptions:

1. census_income_outcome.txt has large data. So, if you want to see the output, change the epoch value to 2 or 3 in NeuralNet.py. As it has many tuples. It takes a lot of time to train the neural net and to get an error of 0 we have to wait for a long time.

2. Last column of the given dataset is considered as the class argument.

3. For standardizing numerical data we use the normalization function -> (x-mean)/(min - max).

4. For n categorical data we divided it into n columns using python library numpy.

5. If any other missing values rather than "?" are used in input data, add it into the miss list in PreProcessing.py

6. If the class value is continuous we classified it based on the mean valued (if the value is > mean then class 1 otherwise,0).

7. The iteration will also be terminated if the specified error tolerance (i.e 0) has been met.

Description about neural network:

1. There is a Neuron class which has layer number, index, bias, adjacencyList, revadjacencyList, net, output.

2. Each nueron except the nerons in the input layer has a bias associated with it(and it is numbered before its nueron).

Analysis:

1. we have attached the output screenshot of the three datasets with our best results

2. for the census_income_outcome.txt there are lot of observations so it takes a lot of time to train the data

3. we have given the output screenshot census_income_outcome.txt for the epoch value of 6 which has taken 8 minutes of time to complete

Log of the Experiments

Car_evaluation	Training Data %	No. of Iterations	Total Training error	Total Test Error	Neurons
1	50	20	0.0446329061775	0.0433106351002	(4,2)
2	**80**	**50**	**0.0151136385231**	**0.0190158195704**	**(5,2)**
3	65	35	0.0442853589217	0.0460188854716	(6,2)

Iris	Training Data %	No. of Iterations	Total Training error	Total Test Error	Neurons
1	**80**	**100**	**0.012168982985**	**0.01635845949**	**(4,2)**
2	80	50	0.0557515652993	0.0490031820439	(4,2)
3	65	35	0.118335907357	0.102784523211	(2,2)

Census_income	Training Data %	No. of Iterations	Total Training error	Total Test Error	Neurons
1	40	3	0.188740691748	0.184588151705	(6,2)
2	60	4	0.119287993413	0.119740003197	(6,6)
3	**80**	**6**	**0.114528927265**	**0.120066365201**	**(8,9)**