Programming Assignment 1: Are we losing customers?[1]

Customer churn occurs when customers stop doing business with a company, also known as customer attrition. It is also referred to as loss of clients or customers.

You are given **sensitive** information of 9,000 of an European Bank, EBQ. Your task is to build an Artificial Neural Network (ANN) based on the dataset such that later the ANN model can predict correctly which of the 1,000 customers (in a separate **unlabeled judge set**) the bank is going to lose. This predictive analysis is vital for the EBQ bank to revise their business strategy towards customer retention. What do you think?

Anyway, you are recruited by the bank to do the analysis. And, the head of the bank only trusts heads, i.e., brains…. I mean neural networks for making any decisions. And luckily you were in Dr. B's class and know something(?) about the ANN that you could successfully convince the head of the bank during the interview that the answer to life, the universe, and everything is the artificial neural network, "and not the number 42". He loved your answer. He put a lot of faith you. Can you solve his problem?

What you are going to submit

A zip file containing the following items:

1. A CSV file (**judge-pred.csv**) containing the predicted labels (by your best performing Artificial Neural Network in terms of accuracy) of the 1000 customers in exact same order of the judge set. Please follow the following format:

   ```
   CustomerID,Exited
   12345,1
   34567,0
   ...
   ```

2. Jupyter notebook (*.ipynb), or the Python (*.py) file(s), with source name "**test-code**" that reads judge.csv file, and generates judge-pred.csv.
3. Jupyter notebook (*.ipynb), or the Python (*.py) file(s), with source name "**training-code**" that builds the artificial neural network of your choice, and reports training and test performance.
4. A Word/PDF file, with name "**Report**" describing the specifications of the artificial neural networks you built with its test performance (on a random 20% test data you set aside from the 9000 customer dataset) in the following (sample) format with imaginary values:

1 Assignment text last update 08/29/18 10:20:11 AM

Report

	Configurations	**Precision**	**Recall**	**F1-score**	**Accuracy**
ANN-1	11x10, ReLU 10x5,ReLU 5x4,ReLU 5x1,Sigmoid Epoch=100 Features=[0:-1] Target=[-1] Encoding={3:'o', 4:'l'} Scaling=[10,11]	Training: 0.99 Test: 0.90	Training: 0.98 Test: 0.85	Training: 0.97 Test: 0.87	Training: 0.99 Test: 0.91
ANN-2	11x6, Sigmoid 6x6,Sigmoid 5x4,Sigmoid 5x1,Sigmoid Epoch=5000 Features=[0:-1] Target=[-1] Encoding={3:'o', 4:'l'} Scaling=[10,11]	Training: 0.99 Test: 0.90	Training: 0.98 Test: 0.85	Training: 0.97 Test: 0.87	Training: 0.99 Test: 0.91
ANN-3	11x10, ReLU 10x5,ReLU 5x4,ReLU 5x1,Sigmoid Epoch=10000 Features=[0:-1] Target=[-1] Encoding={3:'o', 4:'l'} Scaling=[10,11]	Training: 0.99 Test: 0.90	Training: 0.98 Test: 0.85	Training: 0.97 Test: 0.87	Training: 0.99 Test: 0.91
…					
ANN-N	11x10, ReLU 10x5,ReLU 5x4,ReLU 5x1,Sigmoid Epoch=50 Features=[0:-1] Target=[-1] Encoding={3:'o', 4:'l'} Scaling=[10,11]	Training: 0.99 Test: 0.90	Training: 0.98 Test: 0.85	Training: 0.97 Test: 0.87	Training: 0.99 Test: 0.91

Here, ANN-i is the best model showing accuracy 0.99, and we are going to use it to predict the **judge.csv** dataset.

What you are program to accomplish the assignment

- Download the datasets:
  - **dataset.csv** – Containing 9000 customer information (labeled)
  - **judge.csv** – Containing 1000 customer information (unlabeled)
- Preprocess the dataset, and have a 80-20 split (training and test). *Before doing that, please investigate which feature(s) might or might not have relations to the target variable "Exited". You may want to drop those (very) distantly related features from your dataset.*
- Utilize SynapseMatrix style data structure to store weight values and construct your artificial neural networks
- Then, implement backpropagation algorithm.
- Then run the algorithm on the training set, and the apply the trained model on the test dataset.
- Report the structure, algorithm parameters, and performance into the Report document.
- Write a program that has access to your best performing trained ANN (in terms of accuracy metric), and reads the **judge.csv** file. It will output **judge-pred.csv** containing only two fields: CustomerID, Exited for each of the judge instances present in the **judge.csv** file in the exact same order.

CAUTION

- Plagiarism is strictly checked for each submitted codes. Please do not share source files with your fellow classmates. Both parties will be heavily penalized.
- Please do not use Tensorflow, nor Theano, nor Keras to construct and train the ANNs. Using those for this assignment might have grade score 0.
- For every hour after the deadline, 5 points will be deducted, until the score reaches 0.