**STAT 8490 Deep Learning**
**Assignment Four**
**Due by Apr. 1, 6:00pm on Blackboard**

## Problem Description:

In this assignment, we will build an unsupervised self-organizing map (SOM) model to recognize fraudulent credit card transactions. This is very important for credit card companies, because it helps the companies to avoid charging their customers for items that they did not purchase.

The original dataset is available on Kaggle and contains 284,807 transactions made by credit cards in September 2013 by European cardholders. The dataset is highly unbalanced, and the fraudulent transactions only accounts for 0.172%. In order to lighten the computational load, a stratified sample named **creditcard_sampled.csv** is taken from the original dataset and in this assignment, we will use this sample data to build SOMs. The file **creditcard_sampled.csv** is available in our google class folder.

The dataset contains *only* numerical input variables. For confidentiality, these input variables have been transformed into 28 principal components, called V1, V2, ... V28, via Principal Component Analysis (PCA). Only two input variables stay untransformed, 'Time' and 'Amount'. 'Time' gives the seconds elapsed between the current transaction and the first transaction in the original dataset. The variable 'Amount' shows the transaction amount. The dataset also contains a binary variable 'Class' and it reveals the true status of each transaction: 1, if it is a fraud; and 0, otherwise.

## Requirements:

1. Select appropriate input variables (except the 'Class' variable) to build a SOM model. Choose hyperparameters as you feel appropriate.
2. Provide the visualization of the constructed map.
3. Based on the SOM, detect the suspected fraudulent transactions and report the number of detected frauds. Please do not refer to the 'Class' variable, when determining outlying neurons in the map.
4. Compare your fraud predictions from SOM with the actual 'Class' results. Calculate and report the *sensitivity*, percent of true frauds that are correctly identified by SOM.
5. If the model sensitivity is less than 50%, think about approaches of improving the model performance and repeat steps 1-4, until the sensitivity reaches 50% or more.

## Submission Format:

1. Submit your colab notebook called *a4_<yourinitials>.ipynb*. You can decide how much detail of the experiment to include in your notebook. But your notebook should at least include the SOM model of the best sensitivity.
2. Follow the instructions below and convert your colab notebook into a LaTex pdf file called *assignment4.pdf*.

   Finally, make sure that you submit both files (*a4_<your initials>.ipynb* and *assignment4.pdf*) directly to Blackboard. Please do NOT submit a zip file of the required documents.

**To convert Google Colab notebook into .pdf**

1. Download your colab notebook as .ipynb.
2. Upload it to Colab.
3. Run the commands (takes about 1~2 minutes),

*!apt-get install texlive texlive-xetex texlive-latex-extra pandoc*
*!jupyter nbconvert --to pdf  <filename>.ipynb*

4. Refresh the folder.
5. Download the .pdf file to your local machine.