**Introduction to Machine Learning**

**Assignment 1 (20 points)**

# **Logistic Regression:**

We have been talking about logistic regression, one of the most important and fundamental machine learning (ML) algorithms, in the class. It is time for you to execute what you have learned in the class to really implement the logistic regression model. Instead of covering many algorithms quickly at a high-level and using ML algorithm libraries (e.g. [Scikit-learn](http://scikit-learn.org/stable/)), this course aims to explain a few important ML algorithms in-depth. It is meant to prepare you to be a a researcher/scientist — the very top class in the [Machine Learning Skills Pyramid](http://socialmedia-class.org/assets/img/ml_pyramid.jpg), who can not only apply algorithms but create new algorithms!

In this assignment, you will implement logistic regression and get a good understanding of the key components of logistic regression (many other machine learning algorithms as well):

* hypothesis function
* cost function
* decision boundary
* gradient decent algorithm
* gradient checking

You will also apply your implemented logistic regression model to a small dataset and predict whether a student will be admitted to a university. This dataset will allow you to visualize the data and debug more easily. You may find this [**documentation**](http://socialmedia-class.org/slides/AndrewNg_ex2.pdf) very helpful, though it is about how to implement logistic regression in Octave.

#### Instructions

You will use [Jupyter Notebook](http://jupyter.org/), an interactive Python programming and data visualization tool, for this homework. You can follow [this guide](http://jupyter.readthedocs.io/en/latest/install.html) to install Anaconda, that conveniently includes Python, the Jupyter Notebook and other commonly used packages for scientific computing and data science. The starter code is tested for Python 3.6.

You can download the homework assignment zip file from Carmen that contains the starter code, some visualization, and explanatory text. Then, [open the included \*.ipynb file](http://jupyter.readthedocs.io/en/latest/running.html) in the Jupyter Notebook and follow all the instructions to do your work there.

After you are finished, follow the instruction, pack your code into a zip file named like ‘hw2\_yourdotid.zip’, and submit it in the course shell system.

#### Extracurricular

If you are interested in applying logistic regression to some real-world NLP problems, you can try out this task and dataset: [Sentiment Analysis in Twitter](http://alt.qcri.org/semeval2016/task4/). It was an international research competition (so called shared-task). More details can be found in the [overview paper](http://www.aclweb.org/anthology/S16-1001).