# CS-UY/EE-UY 4563: Introduction to Machine Learning

## Overview

This course provides a hands on approach to machine learning and statistical pattern recognition. The course describes fundamental algorithms for linear regression, classification, model selection, support vector machines, neural networks, dimensionality reduction and clustering. The course includes computer exercises on real and synthetic data using current software tools. A number of applications are demonstrated on audio and image processing, text classification, and more. Students should have competency in computer programming

- Prof: Linda Sellie
    - office hours: Tuesday & Thursdays 6:00 - 7`:00 in 10.047 on the 10th floor of 2 MetroTech
- TAs: Shang-Hung Tsai and Ujjwal Singhania
    - office hours: Tuesdays 1:30-3:00 and TBA
- Texts:
    - Learning from Data by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin The book has eChapters which are found at: http://www.amlbook.com/support.html#_echapters)
    - Hands-On Machine Learning with Scikit-Learn and TensorFlow by Aurlien Gron
- Resources:
    - http://cs229.stanford.edu/notes/cs229-notes1.pdf
    - More resources will be provided during the semester
- Supplementary texts and resources
    - Bishop, "Pattern Recognition and Machine Learning"
    - Installing python (need to do this before first recitation): http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html
    - Python tutorial: https://docs.python.org/3/tutorial/
    - James, Witten, Hastie and Tibshirani, "An Introduction to Statistical Learning", https://web.stanford.edu/~hastie/local.ftp/Springer/ISLR_print1.pdf
      Note: While this text uses R, the class will be in Python.
    - Raschka, "Python Machine Learning", 2015. http://file.allitebooks.com/20151017/Python%20Machine%20Learning.pdf

- Grading:
    - Midterm 1: 25%, Midterm 2: 25%, Final project:  20%, Labs, homework & quiz 30%,
    - Labs will involve python-based exercises (most of the labs were developed by Prof. Sundeep Rangan).
    - Final project is done in groups of two.
    - Midterm exams and quiz are closed book.  Students will need to be able to write simple python in the exams.
- Pre-requisites:
    - One of: MA-UY 2224 (Data Analysis); MA-UY 2222 (Data Analysis 2) or EE-UY 2223 (Probability) or equivalent.
    - Undergraduate probability and linear algebra
    - Programming experience is essential, including some exposure or willingness to learn object-oriented programming. No experience in python is required as python will be taught as part of the class.

## Tentative Outline

- Introduction
    - Course logistics. Examples of machine learning problems used today. Formulate machine learning problems (identify task, data, objectives). Classify ML problems as supervised vs. unsupervised, regression vs. classification.
- Linear regression
    - Least squares formula, Gradient Descent, Normal Equations Method
- Model selection and regularization:  Identify the order in a multiple linear regression model
    - Understanding underfitting and overfitting with polynomials; irreducible error; bias and variance tradeoff; cross validation; regularization techniques
    - VC dimension
- Logistic Regression
    - Gradient descent
- Support vector machines (SVMs)
    - Image classification. Support vectors; duality; kernel methods
- Neural networks
    - Formulation; back propagation; Keras
- Convolutional and deep networks
    - Convolutional layers; pooling layers

- PCA
    - Dimensionality reduction
- Clustering and K-Means
    - Unsupervised clustering. K-means; mixture models; EM methods
- (If time allows) Ensemble Learning or Decision Tree Learning
- Final project presentations

## Moses Center Statement of Disability:

If you are student with a disability who is requesting accommodations, please con- tact New York University's Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about CSD can be found at www.nyu.edu/csd. It is located at 726 Broadway on the 2nd floor.

## NYU School of Engineering Policies and Procedures on Academic Misconduct

### Introduction:

The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

### Definition:

Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-

class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.

2. Fabrication: including but not limited to, falsifying experimental data and/or citations.

3. Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.

4. Unauthorized collaboration: working together on work that was meant to be done individually.

5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.

6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

See http://engineering.nyu.edu/files/SACCofC9-14-15.pdf for the Code of Conduct, which can be found here: http://engineering.nyu.edu/life/student-affairs/ community- standards- procedures