IEOR E4525: Machine Learning for OR and FE (Spring 2017)
Syllabus and Course Logistics

Instructors: Martin Haugh and Garud Iyengar
Email: mh2078@columbia.edu and garud@ieor.columbia.edu
URL: `http://www.columbia.edu/~mh2078/` and `http://iyengarlab.ieor.columbia.edu/`

TAs: Octavio Ruiz Lacedelli <or2200@columbia.edu>, Wenjun Wang <ww2438@columbia.edu>, and Shuangyu Wang <sw2756@columbia.edu>

Course Website: All material will be posted on Columbia CourseWorks.

Class Time: Tuesdays and Thursdays 8.40 - 9.55am. Students should **arrive on time** and the use of cell-phones and laptops **will not be permitted** except for running specific course-related applications. Students may be **cold-called** regularly to answer questions in class.

Prerequisites This is intended to be an advanced MS level course for MS students in Operations Research and Financial Engineering. Students should therefore have a good background in optimization, applied probability and simulation. Some familiarity with statistics and in particular, regression and maximum likelihood (ML) techniques, will also be useful. It is also important that students are comfortable with vector and matrix notation and are comfortable with concepts from linear algebra such as the rank of a matrix and the eigen decomposition of a square matrix. Students should also be familiar with at least one of `Matlab` and `R` since we intend to use these software packages / languages extensively throughout the course.

Textbooks: There is no required textbook for **most** of the course as I hope the lecture slides will be sufficient. In addition to the slides, I will also provide lecture notes for a small subset of topics. There are three textbooks that we will regularly use for different parts of the course. They are:

1. *An Introduction to Statistical Learning with Applications in R* (Springer) by James, Witten, Hastie and Tibshirani. This is a very nice textbook and it covers approximately 50% of the material (mainly supervised learning) that we will cover in this course. It is also free to download from the Columbia network at `http://link.springer.com/book/10.1007%2F978-1-4614-7138-7`.

   There will be some assigned readings from this text.

2. *Bayesian Reasoning and Machine Learning* (Cambridge University Press) by David Barber. This is an excellent reference which tends to focus more on Bayesian methods. An electronic version is available at:
   `http://www.cs.ucl.ac.uk/staff/d.barber/brml/`.

3. *Pattern Recognition and Machine Learning* (Springer) by Christopher M. Bishop. This is one of the classic machine learning textbooks.

Other very nice references include:

4. *Learning from Data* (AMLBook) by Abu-Mostafa, Magdon-Ismail and Lin. This is a beautiful book that focuses on the theoretical aspects of learning. While it only considers supervised learning and even then, relatively few algorithms for supervised learning, it is well worth reading and is certainly accessible to advanced undergrads and beginning graduate students.

5. *Mining of Massive Datasets* (Cambridge University Press) by Jure Leskovec, Anand Rajaraman and Jeff Ullman. As the title suggests, this text focuses on problems that are specific to massive data-sets. This is also available online at: `http://www.mmds.org/`.

6. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (Springer) by Hastie, Tibshirani and Friedman. This is also an established machine learning textbook but it reads more like a compendium of various techniques and doesn't have the feel or flow of a textbook. It is also available online at: `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`.

There is also an extensive collection of videos and tutorials on machine learning and data mining that are available online. We may occasionally refer you to some of these additional resources.

Assignments
There will be $n = 9$ or $n = 10$ assignments and students will be asked to complete $n - 2$ of them. Of these n assignments, approximately $m = 5$ of them will be compulsory. Students will then need to complete an additional $n-m-2$ assignments from the remaining $n-m$. Students are welcome to work together on the assignments but each student **must** write up his or her own solution and write their own code. Any student that submits a copy (or partial copy) of another student's solution will receive zero for that assignment and may receive an F grade for the entire course. Late assignments will **not** be accepted!

Exams
The course will have both a mid-term and final exam. Any student who is unable to take an exam must have a very good reason for doing so, e.g., a medical emergency. Such students will take a makeup exam that will be **more difficult** than the regular exam. They will also need to obtain approval from the Dean's office to take such an exam. Exam regrades may be requested by:

1. Explaining in a written statement why you think you should obtain additional points.

2. Submitting this statement and the exam to either the TA or one of the course instructors no later than one week after the exam was returned to the class. (This means that if you failed to collect your exam within a week of it being returned to the class, then you cannot request a regrade!)

It should be kept in mind that when a regrade is requested the entire exam will be regraded and it is possible that your overall score could go down as well as up. **We**

will also photocopy a subset of the exams before returning them to the class. This is intended to deter the very few people (hopefully there are no such people in this class!) who might be tempted to rewrite parts of their exams before requesting a regrade.

Grading
A *tentative* grading scheme is: Assignments 20%, Midterm 35%, Final 45% but I do reserve the right to deviate from this scheme if necessary.

Tentative Syllabus

1. Regression I: linear regression, bias-variance decomposition.
2. Classification I: k-nearest neighbors, the optimal Bayes classifier, naive Bayes, LDA and QDA.
3. Resampling Methods for model assessment and selection: cross-validation methods, the bootstrap.
4. Regression II: subset selection, shrinkage methods including ridge regression and the Lasso, regression in high dimensions.
5. Classification II: Reduced rank LDA, logistic regression, generalized linear models (GLMs), CART, bagging and random forests.
6. A (brief) introduction to causality.
7. Support Vector Machines (SVMs): classification and regression using SVMs, kernel methods and the kernel "trick".
8. The EM Algorithm: applications include clustering via normal mixture models, the general EM algorithm via Kullback-Leibler divergence.
9. Dimension Reduction Methods including principal components analysis (PCA), sparse PCA, kernel PCA, non-negative matrix decomposition, PageRank.
10. Hidden Markov Models (HMMs) including filtering, smoothing and the Viterbi algorithm. Extensions of HMMs.

Our applications will draw from a variety of sources including sentiment analysis, collaborative filtering / recommendation systems, object tracking, social network analysis, web search algorithms, pricing and revenue management, fraud and outlier detection, exploration for natural resources, robotic control, pattern recognition, financial applications and marketing.