**MIDS: W207 Introduction to Machine Learning**

Draft Syllabus (refresh)

Author: Dan Gillick

Last updated: 3/28/2019

**Course Summary**

This course provides a practical introduction to the rapidly growing field of Machine Learning – training *predictive models* to *generalize* to new data. We start with linear and logistic regression, and implement Gradient Descent for these algorithms, the core engine for training. With these key building blocks, we work our way to understanding widely used Neural Network architectures, focusing on intuition and implementation with Tensorflow/Keras. While the course centers on Neural Networks, we will make sure to cover key ideas in Unsupervised Learning and Non-parametric modeling. Along the way, weekly short coding assignments will connect lectures with concrete data and real applications. A more open-ended final project will tie together crucial concepts in experimental design and analysis with models and training.

**Prerequisites**

* Required
  + Core data science courses: W201 and W203
  + Intermediate knowledge of Python (we will use numpy and matplotlib libraries extensively)
  + Basics of Linear Algebra (operations on vectors and matrices)
* Useful but optional
  + Basics of Multivariate Calculus
  + Experience with [Jupyter Notebook](https://jupyter.org/) or [Colab](https://research.google.com/colaboratory/)
  + Experience with Git and [Github](https://github.com/)

**Weekly Schedule**

Topics and optional readings for each week of the course.

1. Introduction and framing
   1. Chollet (1.1, 1.2, 1.3)
   2. RM (1)
2. Linear Regression and Gradient Descent
   1. Chollet (2.2)
   2. RM (10)
3. Feature Engineering
   1. Chollet (5)
   2. RM (4)
4. Logistic Regression
   1. Chollet (3.5)
   2. RM (3)
5. Multiclass classification and metrics
   1. Chollet (6)
   2. RM (3)
6. Feed Forward Neural Networks
   1. Chollet (2.3, 2.4, 2.5)
   2. RM (12, 13)
   3. [Neural Network explainer videos](https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi) (Grant Sanderson)
   4. [Backpropagation derivation](https://www.jeremyjordan.me/neural-networks-training/) (Jeremy Jordan)
   5. [Tensorflow Playground](https://playground.tensorflow.org/) (Daniel Smilkov and Shan Carter)
7. KNN, Decision trees, ensembles
   1. RM (3, 7)
8. Unsupervised learning: k-Means and PCA
   1. RM (11)
9. Embeddings for text
   1. Chollet (4.1, p.524-534)
   2. RM (8)
10. Convolutional Neural Networks
    1. Chollet (8.1, 8.2)
    2. RM (15)
11. Network Architecture Design
    1. Chollet (7.2)
    2. [In-class demo code](https://colab.research.google.com/drive/10jx6M5xO_hgvdqBN-nYK7tJKRM_nPSLZ)
12. Fairness
    1. Essay: [Physiognomy’s New Clothes](https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a) (Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov)
13. Advanced Topics
14. Project Presentations

**Homework Assignments**

In weeks 1-10, students will work (alone) on weekly graded programming assignments that connect concepts in the lectures with concrete tasks

**Final Project**

In weeks 11-14, students will work on a group project (ideally 3-4 people) that is more open-ended. Groups can choose from a few recommended topics (hosted on Kaggle) or potentially select their own. Grading will reflect group size and the application of concepts from the course to the selected task, including data analysis, modeling, experiments, and analysis of errors.

**Grading**

* Assignments (individual work): 65%
* Final project (group work): 30%
* Participation: 5%

Weekly homework assignments can be submitted up to 3 days late with a 10% (absolute) penalty per day. The lowest grade will be dropped.

Baseline grading range for this course is: A for 93 or above, A- for 90 or above, B+ for 87 or above, B for 83 or above, B- for 80 or above, C+ for 77 or above, C for 73 or above, C- for 70 and above, D+ for 67 and above, D for 63 and above, D- for 60 and above, and F for 59 and below.

**References**

No textbook is required, but we suggest some readings from the following sources:

* [François Chollet – Deep Learning with Python](https://search.library.berkeley.edu/discovery/fulldisplay?docid=alma9917107012206531&context=L&vid=01UCS_BER:UCB&lang=en&search_scope=DN_and_CI&adaptor=Local%20Search%20Engine&tab=Default_UCLibrarySearch&query=any,contains,Deep%20Learning%20with%20Python,%20Second%20Edition&offset=0) (available for free for Berkeley students)
* [Dive Into Deep Learning](https://d2l.ai/index.html) (free and open-source)
* [Raschka and Mirjalili -- Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2](https://search.library.berkeley.edu/discovery/fulldisplay?docid=alma9914824770906531&context=L&vid=01UCS_BER:UCB&lang=en&search_scope=DN_and_CI&adaptor=Local%20Search%20Engine&tab=Default_UCLibrarySearch&query=any,contains,Python%20Machine%20Learning:%20Machine%20Learning%20and%20Deep%20Learning%20with%20Python&offset=0) (available for free for Berkeley students)

**Equity and Inclusion**

Integrating a diverse set of experiences is important for a more comprehensive understanding of machine learning. We will make an effort to read papers and hear from a diverse group of practitioners, still, limits exist on this diversity in the field of machine learning. We acknowledge that it is possible that there may be both overt and covert biases in the material due to the lens with which it was created. We would like to nurture a learning environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, veteran status, etc.) in the spirit of the [UC Berkeley Principles of Community.](https://diversity.berkeley.edu/principles-community)

To help accomplish this, please contact your instructor or submit anonymous feedback through ISchool channels if you have any suggestions to improve the quality of the course. If you have a name and/or set of pronouns that you prefer we use, please let your instructor know. If something was said in class (by anyone) or you experience anything that makes you feel uncomfortable, please talk to your instructor about it. If you feel like your performance in the class is being impacted by experiences outside of class, please don’t hesitate to talk with your instructor. We want to be a resource for you. Also, anonymous feedback is always an option, and may lead to your instructor to make a general announcement to the class, if necessary, to address your concerns.

As a participant in teamwork and course discussions, you should also strive to honor the diversity of your classmates.

If you prefer to speak with someone outside of the course, MICS Academic Director Lisa Ho, iSchool Assistant Dean of Academic Programs Catherine Cronquist Browning, and the UC Berkeley Office for Graduate Diversity are excellent resources. Also see the following [link.](https://www.ischool.berkeley.edu/about/community)