**MIDS: Introduction to Machine Learning**
Dan Gillick
Joshua Blumenstock

**Course Summary**
Machine learning is a rapidly growing field at the intersection of computer science and statistics and concerned with finding patterns in data. It is responsible for tremendous advances in technology, from personalized product recommendations to speech recognition in cell phones. The goal of this course is to provide a broad introduction to the key ideas in machine learning. The emphasis will be on intuition and practical examples rather than theoretical results, though some experience with probability, statistics, and linear algebra will be important. Through a variety of lecture examples and programming projects, students will learn how to apply powerful machine-learning techniques to new problems, how to run evaluations and interpret results, and how to think about scaling up from thousands of data points to billions.

**Prerequisites**
1. Students must have completed the following core data science courses prior to enrollment:
    a. Research Design
    b. Storing and Retrieving Data
    c. Exploring and Analyzing Data
2. Undergraduate-level probability and statistics. Linear algebra is recommended.
3. Programming experience, preferably in Python.

**Assignments and Grading**
Course grades will be based mostly on four projects designed to synthesize concepts introduced in the lectures. Each project will require some programming (in Python) to set up experiments and a write-up summarizing results. Grading will be based on correctness and the ability to meet certain experimental milestones (e.g., reducing classification error rate by some percent). Each project will include at least one open-ended question to encourage more research. Course grades will be assessed according to four programming projects and accompanying write-ups.

**Course Resources**
Most textbooks on machine learning are written with considerable technical detail. As a result, there is no one textbook worth following. We will list a variety of readings and web links each week, including some basic science, applications, relevant data sets, and example Python code, when relevant.

**Week 1: Introduction**
- Overview of machine-learning applications

- Brief history
- Fundamentals of machine learning

Read Halevy, A., Norvig, P., & Pereira, F. (2009).The unreasonable effectiveness of data. *Intelligent Systems (IEEE).*
Optional: Provost and Fawcett. *Data science for business*.

**Week 2: Problem Setup and Nearest Neighbors**

- Why prediction?
- Training and test data; cross-validation
- Evaluation and baselines
- Generalization and overfitting: linear models vs. nearest neighbors
- K-nearest neighbors, distance metrics
- Case study: real estate value, digit classification

Read Feynman, R. (1974, June). Cargo cult science. *Engineering and Science 37*(7).
Read Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM.*
Skim Hawkins, D. (2004).The problem of overfitting. *Journal of Chemical Information and Computer Sciences.*

**Week 3: Supervised Learning I: Naive Bayes**

- Probability review: Random variables, Independence, Bayes rule
- Generative models and Naive Bayes
- Maximum likelihood estimation and smoothing
- Case study: spam classification

Read Paul Graham on Naive Bayes in 2002.
Skim Michael Collins tutorial on Naive Bayes (with math), see pages 1–4.
Skim Kanich, C. et al. (2008). Spamalytics: An empirical analysis of spam marketing conversion. ACM conference on Computer and Communications Security.

**Week 4: Supervised Learning II: Decision Trees**

- Decision Trees
- Information Gain
- Overfitting and pruning
- Ensemble methods
- Case study: customer churn, fuel efficiency

Read as much as you can Carter, T. (2001, June). An introduction to information theory and entropy.
Read blog post from yhat about predicting churn.
Read short introduction to Adaboost.

**Week 5: Supervised Learning III: Regression**

- Review of linear regression
- Inference and prediction
- Logistic regression and classification
- Extensions and advanced topics

Read Chapter 5 of Schutt & O'Neill. (2013). *Doing data science*.
Read Section 4.6 of Whitten, Frank, & Hall. *Data mining.*
Optional: Chapter 3 (Sections 3.1 and 3.2), Chapter 4 (especially section 4.4), and Chapter 6 (sections 6.1-6.3) of Friedman, Hastie, & Tibshirani. The elements of statistical learning.

**Week 6: Supervised Learning IV: More Linear Models**

- Gradient descent for regression
- Regularization

Read Chapter 6 of Daum, H. A course in machine learning.

**Week 7: Supervised Learning V: Neural Networks [CSX]: jb**

- The perceptron
- State of the art: neural networks for speech recognition

Read Chapter 8 of Daume. A course in machine learning.
Read Chapter 7 (section 7.4) of Whitten, Frank, & Hall. *Data mining.*

**Week 8: Supervised Learning VI: SVMs, Choosing Classifiers, Speech Recognition**

- Support Vector Machines
- Comparing classifiers: performance, training speed, model size, interpretability
- Feature engineering tips
- Speech recognition overview

Optional Cosma Shalizi SVM lecture notes.
Read An empirical comparison of supervised learning algorithms.
Skim On comparing classifiers: Pitfalls to avoid and a recommended approach.
Skim SKLearn classifier comparisons for toy problems.

**Week 9: Unsupervised Learning I: Cluster Analysis**

- What if our data don't have labels?
- Distance metrics (Hamming, Euclidean, Cosine, Mahalanobis)
- K-means clustering
- Hierarchical clustering

Read Chapter 7 (sections 7.1–7.3) of Rajarman et al. *Mining of massive datasets.*
Read Whitten, Frank, & Hall. Chapter 4.8.
Optional: Zhao, Y., Karypis, G., & Fayyad, U. (2005).Hierarchical clustering algorithms for document datasets. *Data Mining and Knowledge Discovery*.
Optional: Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95: 14863–14868.

**Week 10: Unsupervised Learning II: Expectation Maximization**

- Expectation-Maximization and the idea of hidden variables
- Basics of Gaussian Mixture Models
- Case study: speaker identification

Read Tibshirani lecture notes on EM.
Read Doug Reynolds original paper on GMMs for speaker identification.

**Week 11: Unsupervised Learning III: Dimensionality Reduction**

- Motivation
- Dimensionality reduction
- Principal Component Analysis
- Case study: Eigenfaces
- Other methods for dimensionality reduction: SVD, NNMF, LDA

Read Turk & Pentland. (1991). Eigenfaces for recognition.
Read Chapter 11 (sections 11.1–11.3) of Rajarman et al. *Mining of massive datasets.*
Read Chapter 7 (section 7.4) of Whitten, Frank, & Hall. *Data mining.*
Optional: Chapter 14 (sections 14.2, 14.5–14.10) of Friedman, Hastie, & Tibshirani. *The elements of statistical learning.*

**Week 12: Network Analysis**

Topics

- Graph algorithms (pagerank).
- Network link predictions.
- Scaling and other challenges.

Read Godbole, N. et. al. (2007). Large-scale sentiment analysis for news and blogs. International Conference on Weblogs and Social Media.
Read Page, L. et al. (1999). *The PageRank citation ranking: Bringing order to the web.* Stanford.

**Week 13: Recommender systems**

- Motivation

- The Netflix challenge
- Content-based methods
- Learning features and parameters
- Nearest-neighbor CF

Read Chapter 8 of Schutt & O'Neill. (2013). *Doing data science.*
Read Chapter 9 of Rajarman et al. *Mining of massive datasets*.
Optional: Koren, Y. (2009). The BellKor solution to the Netflix grand prize.
Optional: Resnick et al. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *CSCW*: 175–186.
Optional: Bell, R. M., & Koren Y. (2007). Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter.*

**Week 14: Wrap-Up**
Topics
- Topics beyond the scope of this course
- What your instructors do

Required Readings:
Read Chapter 5 of Whitten, Frank, & Hall.