Learning Bit by Bit
Instructor: Heather Dewey-Hagborg
Office Hours:

From mailing a letter to shopping online to walking down a city street, applications of machine learning have penetrated our daily experience. Our faces, our voices, the emails we write, the products we buy, the content we choose, all constitute our data portrait: aggregates of information that are meticulously sifted, sorted and searched by algorithms behind the scenes.

This class will take a critical tour of the technologies that learn from this data. We will look at the information that defines us and how it is analyzed using techniques common to biology, computer science, robotics and surveillance.

We will cover both the theory and the implementation of machine learning techniques that are commonly used today in applications of text analysis, web search, face recognition, speech recognition, handwriting analysis, and content suggestion. We will discuss the concept of a data portrait and how heuristics and inductive bias shape the way we are seen. Finally, we will apply these techniques to create projects of our own.

This class will involve weekly readings, as well as in and out of class work on individual and group projects engaging with the concepts. Students will be encouraged to implement projects in a variety of media including electronics, robotics, performance, installation, writing, websites, or software.

Prerequisite: H79.2233 Introduction to Computational Media or equivalent programming experience.

Course Goals
Students will learn the concepts behind common machine learning techniques and apply these ideas to projects of their own design.

Expectations
Assignments will include weekly readings and projects. Project mediums will be left open to student interest. Students will be expected to collaborate, to make presentations and to discuss their ideas regularly in class.

Grading
Homework/preparedness 50%
Class Participation 20%
Final Project 30%

Topic Groups Include
1. Content Suggestion - Collaborative Filtering - Clustering
2. Speech Recognition - Natural Language Processing - Markov Models
3. Neural Networks, Supervised and Unsupervised Learning Techniques
4. Face Recognition - Principle Component Analysis

Format
We will spend 3 weeks on each topic grouping. Class time will include a lecture component as well as a lab. There will be weekly technical and critical readings. For each topic students will complete a project either individually, in groups or as a class, depending on student interest. The class will conclude with final project presentations.
Books
Our Readings will come from the following books. All books will be available on reserve in BOBST library. Some books are less expensive and should be purchased if possible. These books are denoted with *
Some books I have found to be fantastic references. These books are denoted with ~. If you are interested in pursuing these topics beyond this class I recommend purchasing one or more of these typically more expensive text books.

--technical
Handbook of Fingerprint Recognition
Author: Davide Maltoni, Dario Maio, Anil K. Jain, Salil Prabhakar
Publisher: Springer, Copyright Date: 2003

*Programming Collective Intelligence
Author: Toby Segaran
Publisher: O'Reilly, Copyright: 2007

*~Natural Language Processing With Python
Author: Ewan Klein, Edward Loper, Steven Bird
Publisher: O'Reilly, Copyright Date: 2009

~Speech and Language Processing
Author: Daniel Jurafsky and James H. Martin
Publisher: Pearson, Prentice Hall, Copyright: 2009, 2nd Edition

~Artificial Intelligence
Author: George F. Luger
Publisher: Addison Wesley, Copyright Date: 2009, 6th Edition
(there is a also a copy of this book in the ITP reading room)

Reliable Face Recognition Methods
Author: Harry Wechsler
Publisher: Springer, Copyright date: 2006

Handbook of Face Recognition
Edited By: Stan Z. Li, Anil K. Jain
Publisher: Springer, Copyright Date: 2005

Introduction to Machine Learning
Author: Ethem Alpaydin
Publisher: MIT Press, Copyright date: 2010

-- critical
*The Numerati
Author: Stephen Baker
Publisher: Houghton Mifflin, Copyright date: 2008

Surveillance: Power, Problems, and Politics
Edited by: Sean P. Hier and Josh Greenberg
Publisher: UBC Press, Copyright Date: 2009
*Niche Envy: Marketing Discrimination in the Digital Age  
Author: Joseph Turow  
Publisher: MIT Press, Copyright Date: 2006

Technologies of Insecurity  
Edited By: Katja Franko Aas, Helene Oppen Gundhus, Heidi Mork Lomell  
Publisher: Routledge-Cavendish, Copyright Date: 2009

~The Philosophy of Mind  
Edited By: Brian Beakley and Peter Ludlow  
Publisher: MIT Press, Copyright Date: 2006, 2nd Edition

*Does Technology Drive History?  
Edited By: Merritt Roe Smith, Leo Marx  
Publisher: MIT Press, Copyright Date: 1994

Classes  
Please bring your laptop to class (if you have one) as we will need them for in-class labs.  
To make lab time as efficient as possible make sure that you have the following software packages installed before class:  
Python 2.6.2  
Java JDK 6 Update 17 or better  
Ant 1.6.0 or better

• Class 1. Introduction  
- Introductions, what we hope to accomplish, go over syllabus  
- Fingerprint Lab

Homework:  
*Collective Intelligence* Ch. 1-2  
(Feel free to skim technical sections on programming with Python if you already feel comfortable)  
*Surveillance: power problems and politics* Ch. “bio-benefits”  
*The Numerati* Ch. “Bloggers”  
Optional: *Handbook of Fingerprint Recognition* Ch. 1

Also, If you don’t have Python installed on your laptop please install by next class.

*Topic 1. Collaborative Filtering*

• Class 2. Collaborative Filtering  
- Discuss readings  
- Lecture 1 on collaborative filtering  
- Students will be assigned to collaborative filtering groups  
- Lab: Filtering the Netflix dataset

Homework:  
- Group Project:
Come up with a group project proposal, begin initial work

*Collective Intelligence* Ch. 3  
*Niche Envy* Ch. 1-3

**Class 3. Collaborative Filtering continued**  
- Discuss readings  
- Lecture 2 on collaborative filtering  
- Lab: in-class time to work on group projects

**Homework:**  
- Group project: Complete projects

**Class 4. Collaborative Filtering continued**  
- Share collaborative filtering projects  
- Lecture 3 on collaborative filtering

**Homework:**  
*Speech And Language Processing* Ch. 1  
*Natural Language Processing With Python* Ch. 1, 2, 5

Follow along with the book installing software and corpora with the examples. Make sure you have NLTK, its dependencies, and the corpora installed before next class. Feel free to skim technical sections on programming with Python if you already feel comfortable.

**Topic 2. Speech and Language**

**Class 5. Natural Language**  
- Discuss readings  
- Introductory lecture on Natural Language Processing.  
- Students will be assigned to speech and language groups  
- Natural Language Lab using NLTK

**Homework:**  
- Group project: Come up with a group project proposal, begin initial work

*Speech And Language Processing* Ch. 9  
*Technologies of (in)security* Ch. “Identification practices”

Also, if you don’t have the latest Java JDK and Ant installed on your laptop please install by next class.  

**Class 6. Speech Recognition**  
- Discuss readings
Introductory Lecture on Speech recognition
Speech Lab using Sphinx4

Homework:
- Group project:
  Complete projects

• Class 7. Projects and advanced topics
  - Share Speech and Language projects
  - Lecture on advanced topics

Homework:
Artificial Intelligence Ch. 11.1 – 11.3 (History, Perceptrons & Backpropagation)
Alan Turing, Computing Machinery and Intelligence:
http://cogprints.org/499/0/turing.html

Topic 3. Neural Networks

• Class 8. Neural Nets
  - Discuss readings
  - Intro lecture on neural nets, perceptrons
  - Neural circuits lab

Homework:
Artificial Intelligence Ch. 11.4 - 11.7 (Competition, Coincidence & Attractors)
The Philosophy of Mind Ch. 78 & 83

- Group project:
  Come up with a group project proposal, begin initial work

• Class 9. Neural Networks continued
  - Discuss readings
  - Intermediate neural networks lecture
  - Lab: work on group projects

Homework:
- Group project:
  Complete projects
  - Watch Ray Kurzweil Video:
    http://mitworld.mit.edu/video/422/

• Class 10. Neural Networks continued
  - Students present neural network projects
  - Discuss Kurzweil video
  - Lecture on advanced topics

Homework:
-Readings for the next class:
How Stuff Works on Facial Recognition:
Look over
http://www.owlnet.rice.edu/~elec301/Projects99/faces/
to prepare for next week’s lab

*Handbook of Face Recognition* Ch. 1
*Does Technology Drive History?* Ch. Technological Determinism in American Culture

-Final projects:
Propose an individual final or facial recognition project idea next week

**Topic 4. Face Recognition**

- **Class 11. Face Recognition**
  - Discuss readings
  - intro lecture on Face Recognition
  - discuss final project ideas
  - lab: ITP faces

**Homework:**
*Introduction to Machine Learning* Ch. 6.1 – 6.5 (Detailed explanation of Principal Component Analysis)
Reading of your choice from *Reliable Face Recognition Methods*
*Does Technology Drive History?* Ch. Do Machines Make History & Ch. Determinism and Pre-Industrial Technology

-Final projects:
Work on final or facial recognition projects

- **Class 12. Face Recognition continued**
  - Discuss readings
  - lecture 2 on face recognition
  - lab: work on projects

**Homework:**
-Final projects:
Work on final or facial recognition projects

- **Class 13. Face Recognition continued**
  - lecture 3 on face recognition
  - Lab: in class discussion about progress and problems and lab on final projects

**Homework:**
-Final projects:
Finish final or facial recognition projects
• Class 14. Final Presentations