

# EECS 445 - Machine Learning

## Lecture 1: WELCOME!!

Date: September 7, 2016

Instructors: Jacob Abernethy and Jia Deng

```
In [10]: from traitlets.config.manager import BaseJSONConfigManager
path = "/Users/jake/.jupyter/nbconfig"
cm = BaseJSONConfigManager(config_dir=path)
cm.update('livereveal', {
    'scroll': True,
    'theme': 'simple',
    'transition': 'fade',
    'start_slideshow_at': 'selected',
})
```

```
Out[10]: {'scroll': True,
'start_slideshow_at': 'selected',
'theme': 'simple',
'transition': 'fade'}
```

## ...Hello World

### Part 1: Administrative stuff

- What is this course?
- Who are we?
- Who should take this course?
- How is the course going to be graded?

### Part 2: Machine Learning? What's that?

- What is ML really?
- Why is it so cool?

# Who are We?

## Professors: (will swap weeks of lectures)

- Jacob Abernethy [jabernet@umich.edu](mailto:jabernet@umich.edu) ([jabernet@umich.edu](mailto:jabernet@umich.edu))
- Jia Deng [jiadeng@umich.edu](mailto:jiadeng@umich.edu) ([jiadeng@umich.edu](mailto:jiadeng@umich.edu))

## Assistants:

- Zhao Fu (GSI) [zhaofu@umich.edu](mailto:zhaofu@umich.edu) ([zhaofu@umich.edu](mailto:zhaofu@umich.edu))
- Chansoo Lee (GSI) [chansool@umich.edu](mailto:chansool@umich.edu) ([chansool@umich.edu](mailto:chansool@umich.edu))
- Benjamin Bray (IA) [benrbray@umich.edu](mailto:benrbray@umich.edu) ([benrbray@umich.edu](mailto:benrbray@umich.edu))
- Valli Chockalingham (IA) [valli@umich.edu](mailto:valli@umich.edu) ([valli@umich.edu](mailto:valli@umich.edu))

## "Prerequisites"

- EECS 492: Introduction to AI
- Undergrad linear algebra (e.g., MATH 217, MATH 417)
- Multivariate calculus
- Undergrad probability and statistics (e.g., EECS 401)
- Programming skills (EECS 280, **EECS 281**) experience in Python
  - Nontrivial level of programming is required.

The only "enforced" prerequisite is 281, but if you are not familiar with either Linear Algebra or Probability/Stat **you are going to struggle in this course**

## This is an UNDERGRADUATE course

### We have received many emails like this:

Dear Sir,

I am joining the ECE Department at the University of Michigan this fall to pursue a Master's degree with specialization in Robotics. I wish to register for EECS 445 (Introduction to Machine Learning) but am unable to do so since I have not completed EECS 281. I am very keen on studying Machine Learning and want to take up the course this semester itself. Would it be possible for you to permit me to enroll for it?)

- Unfortunately, we want EECS 445 to remain an undergraduate focused course
- EECS 545 is meant for graduate students
- Another 545 section was recently opened, to ease the pressure

## Course Grading

- **Homework:** 50% (6 HWs, lowest dropped)
- **Midterm:** 25%
- **Final Exam:** 25%
- Some options for extra credit, details to come

**Canvas Site to be released very soon!**

## Homeworks

- There will be 6 problem sets, roughly one every 2 weeks.
- **Goal:** strengthen the understanding of the fundamental concepts, mathematical formulations, algorithms, and applications
- The problem sets will also include programming assignments to implement algorithms covered in the class.
- Homework #1 will be out next **Monday 9/12** and due following **Friday 9/23**
- Working in groups is fine! You need to report your team members. Limit team size = 4.

## Study Groups

- Form your study group early on!
- Up to four people are allowed.
- For homework, you may discuss between the study group members, but you should write your own solution independently.
- In the homework submissions, you must put the names of other people you collaborated
- Start homework early. (Warning: cramming doesn't work!)

## How to communicate with us?

- No email policy! Instead use Piazza (<https://piazza.com/class/issartijnz3la>)!
- Only exception: personal issues. In this case you can email and/or make an appointment with prof.

## Textbooks

- Much of the material in this course can be learned through online resources
- These two textbooks will be **strongly recommended** although we won't officially designate them as **required**
  - *Chris Bishop*, "Pattern Recognition and Machine Learning". Springer, 2007.
  - *Kevin Murphy*, "Machine Learning, A Probabilistic Perspective". MIT Press, 2012.
- Other recommended texts:
  - *Hastie, Tibshirani, Friedman*, "Elements of Statistical Learning". Springer, 2010. (free online! (<http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf>))
  - *Boyd and Vandenberghe*, "Convex Optimization," Cambridge University Press, 2004. (free online! (<http://stanford.edu/~boyd/cvxbook/>))

## When does this course meet?

### Lectures:

- 001: MW 4:30-6pm (1670 BBB)
- 002: MW 6-7:30pm (Chesebrough Auditorium)

### Discussion Sections:

- 011: F 11:30am-12:30pm (1006 DOW)
- 012: Th 4:30pm-5:30pm (1017 DOW)
- 013: F 1:30pm-2:30pm (1303 EECS)
- 014: Tu 4:30pm-5:30pm (2150 DOW)
- 016: Th 2:30pm-3:30pm (1005 EECS)

**Discussions start TUESDAY next week! No discussion this week!**

# NEW! Sec001 $\neq$ Sec002

## Improving the "multiple course section model"

- Giving a standard lecture back-to-back can be an inefficient use of everyone's time, and don't allow for serious interaction between staff and students
- Repeat lectures unnecessary with video recordings

## We are trying an interesting experiment!

- Each lecture will have **two** versions:
  - A "dry" presentation, with slides and commentary on new material
  - A "hands on" experience, where students work in groups to develop understanding of the material in a structured environment

## How will this work??

- We are *staggering* the lectures in an unusual way.
- Monday 9/12, Section 001, 4:30-6pm: A non-lecture tutorial on python (ignore for now)
- Monday 9/12, Section 002, 6-7:30pm: **Slide presentation of Lecture 02 -- Review of Linear Algebra**
  - Will be video recorded
  - Students are not required to come prepared
- Wednesday 9/14, Section 001: 4:30-6pm: **Hands-on Dive into Lecture 02 material**
  - We will **not** teach Lec02 material
  - Students must arrive **having watchined Lec02 video** or carefully read lecture notes
  - This section will **not** be recorded
- Wednesday 9/14, Section 002: 6-7:30pm: **Slide presentation of Lecture 03 -- Review of Probability/Stats**

## Which should you choose?

- Are you good at preparing before coming to Lecture?
  - Take Sec 001
- Do you prefer to just watch lectures without prep?
  - Take Sec 002
- Do you prefer to watch lectures in your underwear?
  - Great, that's what the lecture video capture is for.

# This course will require you to use Python



- Why is Python a great language for ML?
  - Very simple syntax, code is very concise
  - The libraries are **excellent** and cover a lot of ground (especially for LinAlg, Stats, ML algs)
  - The Jupyter Notebook (<http://jupyter.org>) is a superb tool for communicating data analysis and visualization

## Jupyter Notebook? What's that?



- Interacting with Python via Jupyter is Awesome!
- "Jupyter" formerly known as "IPython Notebook"
- This lecture (and many to come) is actually a Jupyter Notebook!
- Easy to display code, code output, rich text, mathematics (via latex/mathjax), all within the same document

```
In [5]: x = 2
        x = x * 2
        print("Here is some math: %d + %d = %d" % (x, x, x + x))
        print(" how are you??")
```

```
Here is some math: 4 + 4 = 8
 how are you??
```

## Python: We recommend Anaconda (Python 3.5 suggested)



- Anaconda is standalone Python distribution that includes all the most important scientific packages: *numpy, scipy, matplotlib, sympy, sklearn, etc.*
- Easy to install, available for OS X, Windows, Linux.
- Small warning: it's kind of large (250MB)

## Some notes on using Python

- HW1 will have a very simple programming exercise, just as a warmup.
- This is a good time to start learning Python basics
- There are **a ton** of good places on the web to learn python, we'll post some
- This course requires you to pick up skills in python on your own, we won't devote much lecture time!
- We may require some homeworks to be submitted in the jupyter notebook format.

## Checking if all is installed, and HelloWorld

- If you got everything installed, this should run:

```
# numpy crucial for vectors, matrices, etc.
import numpy as np
# Lots of cool plotting tools w/ matplotlib
%pylab inline
# For later: scipy has a ton of stats tools
import scipy as sp
# For later: sklearn has many standard ML algs
import sklearn as skl
# Here we go!
print("Hello World!")
```

## More on learning python

- We will have one tutorial devoted to this: Monday's hands-on lecture (4:30-6pm)
- If you're new to Python, go slow!
  - First learn the basics (lists, dicts, for loops, etc.)
  - Then spend a couple days playing with numpy
  - Then explore matplotlib
  - etc.
- Piazza = your friend. You can ask anything you like about using Python etc.

## Pitch: Join the Michigan Data Science Team!



- Started by student Jonathan Stroud and Jake a year ago
- Hack on data science and ML challenges, win prizes!
- We've gotten some serious attention for our work on the Flint Water Crisis ([Gizmodo](http://gizmodo.com/google-is-helping-flint-prioritize-which-lead-pipes-nee-1774517926) (<http://gizmodo.com/google-is-helping-flint-prioritize-which-lead-pipes-nee-1774517926>), [Chicago Tribune](http://www.chicagotribune.com/bluesky/originals/ct-google-flint-water-data-bsi-20160503-story.html) (<http://www.chicagotribune.com/bluesky/originals/ct-google-flint-water-data-bsi-20160503-story.html>), [Detroit Free Press](http://www.freep.com/story/news/local/michigan/flint-water-crisis/2016/05/03/google-u-m-partner-help-flint-water-crisis/83836678/) (<http://www.freep.com/story/news/local/michigan/flint-water-crisis/2016/05/03/google-u-m-partner-help-flint-water-crisis/83836678/>))
- Infosession for [MDST](http://mdst.eecs.umich.edu/) (<http://mdst.eecs.umich.edu/>) is Thursday (tmw!) at 6pm in 1670BBB

## Welcome aboard everyone!

- We're going to have a lot of fun this semester
- We also want to hear your feedback, so feel free to share using Piazza private posts
- Now for Jia's portion...!