Machine Learning 08/22/2016
Bramson Notes

Tues/Thurs 1pm = office hours
Each class has a jupyter notebook

What is Machine Learning (ML)?

- Make machines think like humans. 1st thought in ~1980s was that distinguishing pictures of cat vs. dog would be easy- that a couple undergrads could do it in a summer and that we'd have something like Google translate by the 80s
    - Lolz, we're still struggling with it
    - Overpromises led to losing public trust and funding -> led to the "first winter of AI" a renaissance of sorts" and AI (artificial intelligence) was rebranded as Machine Learning
    - As we get more data, information -> use AI techniques (but people weren't calling them this)
    - Now is a field that is growing faster than other disciplines- +1000 people/year at AI/ML conferences

Showcases of ML's increasing use/popularity

- Go- a Japanese game
    - no 2 games are like each other – different than Chess, as chess has limited possibilities (pieces with set moves/options); Go has exponential branching at each move – so it was thought ML would not work very well in this case
    - BUT, earlier this year, ML beat the world champion for the first time – this was surprising as people didn't think ML could "learn" such an intuitive game
- Google just bought a ML company

Unspoken objective of this class = give set of tools to increase possibilities after grad school

A few case studies of ML:

- Mapping before/after images of flood uses ICA (an unsupervised learning technique). Can use it to go to a new image, run the technique and then have a good approximation of the extent of new flood.
- Hard to track povery – need people to do census but that works only if have people to go door to door, and if people actually live in the houses the census takers visit. Also not reasonable to track poverty month-to-month or year-to-year.
    - Can use Landsat images with ML (convolution neural nets) to predict and track poverty: red = high probability of poverty in Uganda
    - It learns roads in bad condition = correlated with poverty

  - Question from class: what about the lobe in upper right that is red, but doesn't match 2005 poverty map. Leon = that is area where no data so it just filled in with the high probability results from around it
- TARS (traverse Aeolian ridges) are small dunes on Mars. ML algorithm takes image and assigns probability of dune pixel by pixel. This would be intractable if done by human. Generally when people do mapping, can only consider bigger areas, not pixel-wise.

This class is Machine Learning/Data Science
Jupyter = good way to do tutorials with text and code + can run code in browser
We're going to use python – newest ML in python environment
2 ways to do python

1. "drive stick" – use version 2.7 since most programs support – this installs it all
2. "drive automatic?" – install Canopy – a python version of matlab; a simple environment where you can easily install packages

   We'll use packages: pandas (data processing), matplotlib (to plot), numpy, scipi and scikit (python's widest used ML tools)

Not an assignment but you should all get this all installed and running by next class

Python overview: you can assign variables, make matrices, print var to see values
Math package – don't use it as crashes with numpy – uses different data types

Plotting in python- plt.show() to show plots

Pandas – amazing way to organize data – 1 line to import csv files, loads into data frame (an aside: R also uses data frames and matlab has also started implementing them)
With Pandas you can access each by name – and empty values are assigned NaN
..dropna() to drop NaNs, .notnull to see if a NaN
3 lines of code to clear NaNs from part of a table

Can make histograms: .value_counts()

Jess question about "do not use in canopy" comment in the jupyter workbook:
In jupyter: plt.show() doesn't work but need to do a certain way (that's the line in the jupyter document that says to not use in in canopy