Course Admin

EE-UY 4563/EL-GY 9123: INTRODUCTION TO MACHINE LEARNING

PROF. SUNDEEP RANGAN

Course Details

- Prof: Sundeep Rangan, 2 MTC, 9.104, srangan@nyu.edu
  - Office hours: Wednesdays 9-11
- TAs:
  - Ish Kumar Jain, EE-UY 4563, ishjain@nyu.edu
  - Mengzhe Huang, EL-GY 9123, m.huang@nyu.edu
- Graders: Large team.
- Lectures:
  - EE-UY 4563 TuTh 4:30-5:50 RH 315
  - EL-GY 9123 Tu 6:30-9 Pfizer Auditorium

Grad vs Undergrad

- Class is simultaneously offered at the graduate and undergraduate level
- Undergrad EE-UY/CSE-UY 4563: Intro to Machine Learning
  - Covers fundamental algorithms and some analysis
  - In depth coverage of software tools including python, Google Cloud, Tensorflow
  - Python-based lab exercises + mandatory project
- Grad EL 9123: Intro to Machine Learning
  - More algorithms and more mathematical analysis. Faster paced.
  - Software tools must be learned at home. Less coverage in class
  - Python-based lab exercises + optional project
- Lecture notes are mostly common with supplementary material for grad students indicated
- Many labs are common

Texts and Other Resources

- Undergrad: James, Witten, Hastie and Tibshirani, "An Introduction to Statistical Learning",
  - http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf
  - Very clear explanation of concepts.
  - But examples are in R. And there is no review of probability
- Grad: Hastie, Tibshirani, Friedman, "Elements of Statistical Learning"
  - https://web.stanford.edu/~hastie/Papers/ESLII.pdf
  - More advanced text with more analysis
- Raschka, "Python Machine Learning", 2015.
  - http://file.allitebooks.com/20151017/Python%20Machine%20Learning.pdf
  - Excellent examples of using Python
- Bishop, "Pattern Recognition and Machine Learning" (more advanced)
- Coursera course: Generally do not cover probability
- Undergrad probability

Pre-Requisites

- Undergrad probability required for both UG and Grad version:
  - Basics of random variables, densities, Gaussian distributions, correlation, expectation, conditional densities, Bayes' theorem
  - Will provide a short review
  - NYU classes: Data analysis or Intro Probability are sufficient
- Calculus and Linear algebra
  - Vectors, matrices, partial derivatives, gradients.
  - Undergrad class will provide a brief review
- No machine learning experience is necessary
  - If you have ML experience, do NOT take this class.
  - Take Graduate probability (Fall) then Advanced machine learning (Spring)

Pre-Requisites Programming

- Python
  - All labs are in python, similar to object-oriented MATLAB, but many more libraries.
  - And free!
- What you need to know
  - You do not need to know python before class. But, we will go over it quickly.
  - You should have experience in some programming language (eg. MATLAB).
  - You should know or being willing to learn object oriented programming
- Resources:
  - Installing python and ipython notebook (make sure you install Version 3.5) http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html
  - Python tutorial: https://docs.python.org/3/tutorial/
  - Numpy: http://cs231n.github.io/python-numpy-tutorial/

Grading: Undergraduate

- Midterm 1: 20%, Midterm 2: 20%, Final: 20%, Labs, HW and quizzes: 20%, Final project: 20%
- Labs: Simple python exercises
  - Given as ipython notebook that you complete.
- Midterms & final
  - Each over approx. 3-4 weeks of material
  - Closed book with cheat sheet.
  - Follows homework and quiz problems + some very basic python questions
- Final project:
  - Use machine learning in some interesting way.
  - Must use data and python analysis.
  - Provide final report.

Grading: Graduate

- Midterm 30%, Final 30%, Labs / HW 30%, Quizzes 10%
- Optional project: Up to 20%
- Labs: Simple python exercises
  - Given as ipython notebook that you complete.
- Midterms & final
  - Each over approx. 6-7 weeks
  - Open book but no electronic aids.
  - Follows homework and quiz problems + some very basic python questions
- Optional final project:
  - Use machine learning in some interesting way.
  - Must use data and python analysis.
  - Provide final report.

Learning Objectives

- ❑ Formulate a problem as a machine learning problem
  - Identify learning objectives, source of data, models, …
- ❑ Load data from various source
- ❑ Visualize data
- ❑ Mathematically describe simple models of the data
- ❑ Fit the models to data and use models for prediction, estimation, …
- ❑ Evaluate the performance of methods using statistical techniques

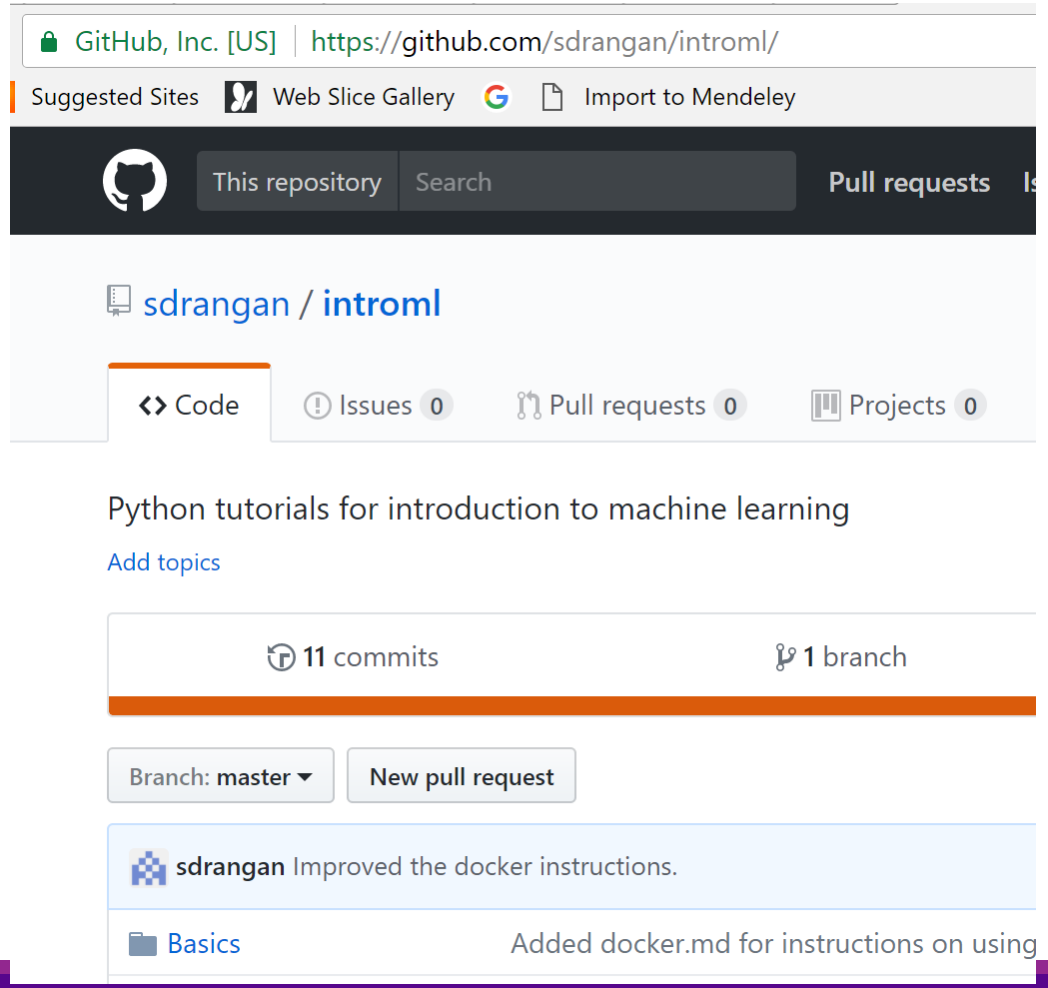
Machine Learning Project

- ❑ Perform an interesting machine learning task of your choice
- ❑ Many possible areas:
  - Machine vision, brain-computer interfaces, natural language processing, sentiment analysis, …
  - Anything that interests you
- ❑ Use real data
  - UCI ML repository
  - Google BigQuery data
- ❑ Write code
- ❑ Submit report in a conference format
- ❑ Poster presentation at end of class
- ❑ 20% of grade

Github

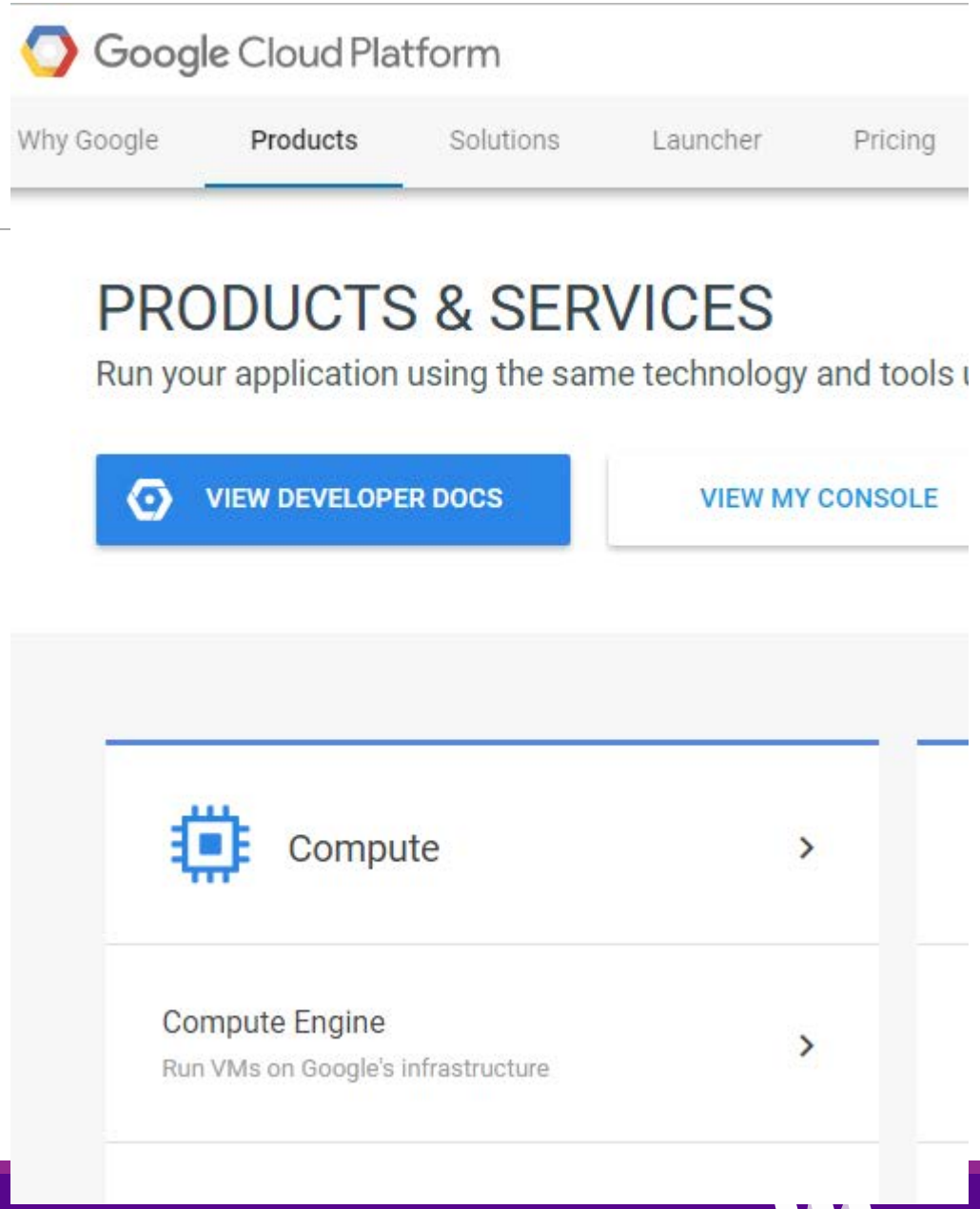
- Labs and demo posted on github
- https://github.com/sdrangan/introml/
- Also includes instruction for installing software

- Several tutorials of github on the web.
- Available on Windows, Mac and Unix.
- But, you can just clone the repo



Google Cloud Platform

- All labs in this class can be run on either:
  - Your own computer: Windows, MAC
  - Google Cloud Platform (GCP)
- GCP pros and cons:
  - Access to powerful machines / large storage for projects. Includes GPUs
  - Access to many services such as BigQuery
  - Can scale your computational resources
  - But, somewhat harder to sync editors / debuggers
- Getting started: https://cloud.google.com/
- Instructions on https://github.com/sdrangan/introml/tree/master/GCP



Other Software

- On your machine (local or GCP), you will need to install several pieces of software:
- Python with various packages
  - Make sure you get 3.5
  - Anaconda
  - Jupyter notebook
  - See notes in NYU Classes
- Tensorflow (needed only later in the class)
- Git hub
  - Guides: https://guides.github.com/
  - Available on Windows, Mac or Linux (including GCP instances)
  - All demos will be available on: https://github.com/sdrangan/introml.git