Data-Science Boot-camp

January 9th – 13th, 2017
10am to 4pm (9am to 10am, Introduction January 9th)
524 S.W. Mudd Building

The Collaboratory@Columbia is pleased to announce a free Winter Boot-camp for Columbia University PhD students and postdoctoral scholars who are interested in extending their existing mathematical and programming skills to include a training in data science. Designed by faculty and postdoctoral scholars from Columbia University's Data Science Institute, the curriculum includes on-line learning material, introductory lectures, hands on laboratory experiences and a capstone project. Click here for detailed course information, including basic prerequisites. Click here to apply to participate. The deadline for applications is January 2nd, 2017.

Jointly founded by Columbia University's Data Science Institute and Columbia Entrepreneurship, **The Collaboratory@Columbia** is a university-wide program dedicated to supporting collaborative curricula innovations designed to ensure that all Columbia University students receive the education and training that they need to succeed in today's data rich world.

Course Information

The course is a blend of online learning experiences (about 2 hours of preparation will be required per day), in-class lectures, hands-on laboratory exercises with a variety of data sets, and a capstone project. The course will use the Python programming language. Participants are required to bring their own computer to the daily sessions. Lunch and refreshments are not provided.

Day 1: January 9th, 2017: INTRODUCTION TO DATA SCIENCE
Introduction to Data Science; Data Visualization; Probability and Regression

- 9am-10am: Welcome and Registration
- 10am-12pm: Session 1 (1h introduction + 1h lab)
- 12pm-1pm: Lunch Break
- 1pm-2:30pm: Session 2 (30 min introduction + 1h lab)
- 2:30pm-3:00pm: Break
- 3:00pm-4:00pm: Session 3 (1h lab)

Day 2: January 10th, 2017: ALGORITHMS & CLASSIFICATION
Introduction to Algorithms; Introduction to Machine Learning; Classification

10am-12pm: Session 1 (1h introduction + 1h lab)
12pm-1pm: Lunch Break
1pm-2:30pm: Session 2 (30 min introduction + 1h lab)
2:30pm-3:00pm: Break
3:00pm-4:00pm: Session 3 (1h lab)

Day 3: January 11th, 2017: MACHINE LEARNING
Model Selection; Probabilistic Modeling; Evaluation

10am-12pm: Session 1 (1h introduction + 1h lab)
12pm-1pm: Lunch Break
1pm-2:30pm: Session 2 (30 min introduction + 1h lab)
2:30pm-3:00pm: Break
3:00pm-4:00pm: Session 3 (1h lab)

Day 4: January 12th, 2017: ADVANCED TOPICS IN DATA SCIENCE
Natural Language Processing; Information Retrieval; Neural Networks

10am-12pm: Session 1 (1h introduction + 1h lab)
12pm-1pm: Lunch Break
1pm-2:30pm: Session 2 (30 min introduction + 1h lab)
2:30pm-3:00pm: Break
3:00pm-4:00pm: Session 3 (1h lab)

Day 5: January 13th, 2017: CAPSTONE PROJECT

10am-12pm: Session 1 (Lab Project Work)
12pm-1pm: Lunch Break
1pm-2:30pm: Session 2
2:30pm-3:00pm: Break
3:00pm-4:00pm: Session 3 (Lab Project Work)

Prerequisites

Basics of linear algebra
Basics of statistics (mean, variance, etc.)
Basic programming skills in Python
Basic understanding of data structures and algorithms
Basic skills for working with data files (i/o operations on csv and tsv files, etc.)

To get started here are some online resources for programming skills in Python and for setting up the required Python development environment.

Coursera "Programming for Everybody (Getting Started with Python)":

- https://www.coursera.org/learn/python

We'll be using the Anaconda Python distribution:

- http://docs.continuum.io/anaconda/faq.html

Along with the Jupyter notebook with an IPython kernel:

- https://jupyter.readthedocs.io/en/latest/content-quickstart.html
- http://opentechschool.github.io/python-data-intro/core/notebook.html
- http://nbviewer.ipython.org/github/catherinedevlin/mpwfw_exercises/blob/master/setup.windows.ipynb