Course Introduction

Communication

- Communication links:
  - maria.hybinette AT mac.com
- Piazza
  - Setting up
- Access to Server to test and run your programs
  - Setting up.

Grading & Class Work


Course Topic

- What is “Machine Learning for Trading”
  - Applies machine learning strategies to real world trading decisions.
    - We will utilize real world stock data
    - We will program in python
  - Audience: All majors, not just computer science majors, still assumes programming skills
Course Topic

• How Does it Differ from:
  – CS 4641: Machine Learning?
    • Our (4645) course is an applied course, we will learn machine learning by programming in python and its modules.
    • Learn by example.
  – CS 7646: On-line version of the class?
    • Graduate version, self-directed course.
    • No class interaction.
    • Similar content.

3 Parts of Course

1. Real World Data: Manipulating Financial Data in Python
   – Read historical financial data into python and manipulate it using powerful statistical algorithms

2. Real World Strategies: Computational Investing
   – Algorithms, methods and models used by hedge funds and investment banks to manipulate and work with financial data

3. Add Learning to (1)+(2): Learning Algorithms for Trading
   – We pull (1) and (2) together:
     • Take what we learned in the first two segments:
       – Data manipulation and
         – Classic investment strategies in the real world and
     • Show how to take that data and use it with learning, machine learning, like Q learning and random forests to build new trading algorithms

Text Books

• "Python for Finance: Analyze Big Financial Data", Yves Hilpisch
  – Chapters 4, 5, 6, 11
• "Machine Learning", Tom M. Mitchell
  – Chapters 1, 3, 8, 13
• "What Hedge Funds Really Do", Philip Romero and Tucker Balch
  – Chapters 2, 4, 5, 7, 8, 9, 12

Prerequisites

• Strong programming skills!
  – Main requirement.
• Some python experience
• Install python (+ numpy, scipy, pandas, matplotlib) framework on laptop that is brought into class every lecture
  – Will use for ‘activities’
Why Python?

• Quick Prototyping
• Easy
  – to Learn
  – to Use
  – to Read – reads like English
• Document rich
• Intuitive guess what should work, and it works.
• Powerful Libraries or Modules
  – 3rd party: Example numPY

Who uses Python?

• United Space Agency - NASA
• Google: Maps, Gmail, Groups, News
• YouTube, Reddit, BitTorrent
• Computational Finance
• Research: Universities worldwide for a variety of disciplines

Python Primers

• We will cover the highlights of python related to computational finance.
  – We will provide ‘templates’ on what you need related to topics covered in class.
    • We do assume strong programming skills and motivation.
  – If you want dwell deeper:
• Good resources:
  – “Dive into Python”, Mark Pilgrim
    • http://diveintopython.net
  – The Official Python Tutorial
    • https://docs.python.org/3/tutorial/
  – The Python Quick Reference:
    • http://rgruet.free.fr/#QuickRef

Overview of Libraries Module (we will learn how to use these by example)

• NumPy –
  – Numerical python, array oriented programming
  – Provides powerful data structures for efficient (memory) computation (operations) of arrays and multi-dimensional arrays and matrices.
• SciPy
  – Extends NumPy: Adds scientific algorithms:
    • integration, interpolation, minimization, regression, linear algebra, and statistics.
• Pandas
  – Spreadsheets for python
  – Good for analyzing tabular data (likes spread sheet data)
  – Structured data operations and manipulations
  – Data Frame.
• Matplot lib
  – Plotting mostly 2D, some limited 3D plotting is available.