Python for Analytics – Syllabus

Summer – 2018

Course Description

This course in python starts with introduction to the python programming language basic syntax and environment. It methodically builds up the learner's experience from the level of simple python statements and expressions to writing succinct, efficient and fast Python expressions and package the code in methods and classes.

In general, the course is geared toward developing a data science's toolbox such as data importing, cleaning and preparation and covers a number of machine learning algorithms. However the course expands beyond these skills as it stresses upon the importance of some of Python's most and unique and powerful features and serves as an introduction to object oriented programming and Python Classes.

Instructor

Hussam Almuayad (ha01@UChicago.edu)

Learning Objectives

The course will enable students to:

- Understand the architecture the Python language such as module, Classes, Class instances and Classs methods and instance methods.
- Establish a solid understanding of Python programming language, including Python containers, expressions, statements, comprehension, and user defined functions.
- Have hands on experience using a number of core modules essential for data science applications, pipelines, regular expression and text processing as well as statistical analysis.
- Import, manipulate, index, clean and export datasets.
- Perform exploratory data analysis and simple model estimation and apply a number of machine learning techniques.
- Understand how to use rest APIs to extract data from the web.

What this course is Not

This course is not an introduction to the theories underlying the different techniques in machine learning. We will touch upon a number of examples that discuss different modeling and ML techniques; however, the examples aim to demonstrate the methods of applying the techniques in Python.

Tools

Anaconda (latest with Python v3.6). Download here
IPython (Jupyter) Notebook
Packages used: sys, os, re, numpy, pandas, matplotlib, scikit-learn, and statsmodels.

Course Format

The course is comprised of series of Jupiter notebooks demonstrating and explaining the different topics covered in an increasing level of difficulty building on previous topics. Each Notebook contains missing code blocks intended to be completed during the session.

Textbooks and Resources

The course is a distillation of many books and online courses, most notably:

Learning Python 5th Edition, Mark Lutz
Learn Python 3 the hard way, Zed Shaw
Python for Data Analysis, Wes McKinney

Course Schedule

Lecture 1:

Anaconda Installation, creating multiple conda environments and using IPython Notebook.
A Brief history of Python. The mental shift and Why use python?
Python core object types, File input/output, and string fundamentals

Lecture 2:

Conditional statement in Python
User defined functions, python comprehensions, lambda functions exception handling

Lecture 3:

Handling date and time in Python
Saving structured Python objects
regular expression

Lecture 4:

[Numpy] Datatypes, structures, manipulating and slicing, arithmetic operations
[Pandas] DataFrames and Series, structures, data subsetting, indexing, summary statistics, and integration with numpy.

Lecture 5:

[Pandas] Continued
Introduction to plotting in Python using Matplotlib and Seaborn.
Data sources and using Python ODBC.

Lecture 6:

Download a dataset using FTP, how to prepare and clean a messy data set.
Time series analysis using statsmodels

Lecture 7:

Data imputation using pandas, and Scikit-learn.
Statistical modules: Scikit-learn and statsmodels for OLS and linear regression.

Lecture 8:

Logistic Regression and shrinkage in Python
Classification Trees and Random Forests Classifier

Lecture 9:

Regression Trees and Random Forest Regressor
Python Pipelines

Lecture 10:

Python Classes, instances of classes and object inheritance.
Understand python Classes by building a generic example
Created a Python Class that performs Linear regression.