2020 Python Boot Camp Syllabus

Educational Aim

The goal of our Python Boot Camp is to provide you with the skills needed to produce a portfolio worthy data science/machine learning project.

1 Topics to be Covered

In fitting with our educational aim, this course will walk you through the steps involved in a typical data science/machine learning project. While covering a variety of important techniques and algorithms we will illuminate themes and motivations that can be adapted to many data science settings.

Below we outline the specific topics we will be covering in this boot camp.

- Data Gathering Techniques
    - Searching common online sources for data
    - Basic web scraping with BeautifulSoup
    - Interacting with APIs

- Data Cleaning
    - Data Types
    - Basic data exploration with pandas, and numpy
    - Basic plotting with matplotlib
    - Handling Missing Data
    - Common Data Transformations

- Supervised Learning
    - Regression
        * Simple Linear Regression
        * Multiple Linear Regression
        * Polynomial Regression
        * Ridge Regression
        * LASSO
        * Kernel Regression (if time permits)
        * Local Regression (if time permits)
    - Classification
        * Nearest Neighbor Methods

- Naive Bayes
- Logistic Regression
- Decision Trees
- Random Forests
- Support Vector Machines

- Unsupervised Learning

  - Dimensionality Reduction
    * Principal Components Analysis
    * t-SNE
  - Clustering
    * k-Means
    * Hierarchical
    * DBScan

- Forecasting for Time Series Data

  - Handling and cleaning time series data
  - Simple forecasting methods
  - Time series regression models
  - Smoothing
  - Exponential Smoothing

- Neural Networks

  - Perceptrons
  - Shallow Networks

- Presenting Results

  - Pandas for presentation
  - Advanced matplotlib
  - Plotting in seaborn
  - Introduction to Interactive Plotting With Python

- Machine Learning Concepts

  - Training Test Split
  - Loss Functions
  - Gradient Descent
  - Model Validation
  - Bias Variance Trade-Off
  - Cross Validation

Course Structure

The boot camp will meet for twelve one and a half hour long sessions that are led by an instructor. Most lectures will be accompanied by a homework assignment covering that lecture's materials. At the end of the twelve lectures students will be given a week to work on group projects. During that week lectures will be replaced with open office hours in which lecturers and mentors may provide guidance for the group projects. The boot camp will end with a presentation day for all group projects.

Lectures

Each lecture the instructor will progress through a series of jupyter notebooks as well as prepared python code. These sessions will feature a blend of lecturing and working in small groups through coding examples. Students are encouraged to ask questions during lectures in order to ensure they understand the material.

Homework

Most sessions will have a corresponding homework set that will highlight the material covered in said session. Completion of homework sets is not mandatory, but highly encouraged. Working through problems and examples on your own will enrich your experience in the boot camp.

Group Projects

Each participant must complete a group project by the end of the bootcamp. Projects should feature the data science/machine learning skills taught in the bootcamp, or even feature more advanced techniques groups teach themselves outside of bootcamp.

References

All lectures and homeworks draw upon the material presented in the following books:

- Python for Data Analysis
- Data Science from Scratch
- Introduction to Machine Learning with Python
- An Introduction to Statistical Learning
- The Elements of Statistical Learning
- Hands-On Machine Learning with Scikit-Learn and TensorFlow
- Hands-On Unsupervised Learning Using Python

- The Hundred-Page Machine Learning Book

- Forecasting: Principles and Practice

- An Introduction to Neural Networks