Data Science Boot Camp

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **5 Day**

https://bilginc.com/en/training/data-science-boot-camp-40-training/

Overview

Introductory Python (1 day)

This is a class for computer-literate people with no programming background who wish to learn basic Python programming. The course is aimed at those who want to learn "data wrangling" – manipulating downloaded files to make them amenable to analysis. We concentrate on language basics such as list and string manipulation, control structures, simple data analysis packages, and introduce modules for downloading data from the web.

This Introductory Python class is designed for computer-literate people with no programming background who wish to learn basic Python programming. The course is aimed at those who want to learn "data wrangling" – manipulating downloaded files to make them amenable to analysis. We concentrate on language basics such as list and string manipulation, control structures, simple data analysis packages, and introduce modules for downloading data from the web.

2. Data Science with Python: Data Analysis and Visualization (2 days)

Beginner

This class is a comprehensive introduction to data science with Python programming language. This class targets people who have some basic knowledge of programming and want to take it to the next level. It introduces how to work with different data structures in Python and covers the most popular data analytics and visualization modules, including numpy, scipy, pandas, matplotlib, and seaborn. We use Ipython notebook to demonstrate the results of codes and change codes interactively throughout the class.

Overview

This class is a comprehensive introduction to Python for Data Analysis and Visualization. This class targets people who have some basic knowledge of programming and want to take it to the next level. It introduces how to work with different data structures in Python and covers the most popular Python data analysis and visualization modules, including numpy, scipy, pandas, matplotlib, and seaborn. We use Ipython notebook to demonstrate the results of codes and change codes interactively throughout the class.

Prerequisites

If you have good knowledge of basic data types (e.g. string, numeric), data structures (e.g. list, tuple, dictionary) and are familiar with concepts of list comprehension and for/while loop, you are good to go with the Python for Data Analysis and Visualization course. We will cover these basic Python programming topics in the course as well, but move at a relatively fast speed.

3. Big Data with Hadoop and Spark (2 days)

Course Overview

This program providing a hands-on introduction to the Hadoop and Spark ecosystem of Big Data technologies. The course will cover these key components of Apache Hadoop: HDFS, MapReduce with streaming, Hive, and Spark. Programming will be done in Python. The course will begin with a review of Python concepts needed for our examples. The course format is interactive. Students will need to bring laptops to class. We will do our work on AWS (Amazon Web Services); instructions will be provided ahead of time on how to connect to AWS and obtain an account.

This program providing a hands-on introduction to the Hadoop and Spark ecosystem of Big Data technologies. The course will cover these key components of Apache Hadoop: HDFS, MapReduce with streaming, Hive, and Spark. Programming will be done in Python. The course will begin with a review of Python concepts needed for our examples. The course format is interactive. Students will need to bring laptops to class. We will do our work on AWS (Amazon Web Services); instructions will be provided ahead of time on how to connect to AWS and obtain an account.

Prerequisites

To get the most out of the class, you need to be familiar with Linux file systems, Linux command line interface (CLI) and the basic linux commands such as cd, ls, cp, etc. You also need to have basic programming skills in Python, and are comfortable with functional programming style, for example, how to use map() function to split a list of strings into a nested list. Object oriented programming (OOP) in python is not required.

Outline

Introductory Python (1 gün)

1. Ünite: Liste düzenleme

- Basit değerler ve ifadeler
- Fonksiyonları tanımlama, olağan söz dizimlerini ve Lambda söz dizimlerini kullanma
- Listeler
- Bütünleşik fonksiyonlar ve alt simgeleme
- İçiçe geçmiş listeler
- Fonksiyonel operatörler: eşleştirme ve filtreleme
- Listeleri kavrama
- Çoklu liste işlemleri: eşleştirme ve sıkıştırma
- Fonksiyonel operatörler: azaltma

2. Ünite: Diziler ve basit I/O

- Karakterler
- Karakter listeleri olarak diziler
- Bütünleşik dizi işlemleri
- Dizi listeleri olarak girdi dosyaları
- Baskı tümcesi
- Web'den veri okuma
- İstek paketini kullanma
- Dizi tabanlı web düzenleme (örn. csv dosyalarını ele alma)

3. Ünite: Kontrol yapıları

- Tümcelere karşı ifadeler
- Döngüler için
- Döngülerdeki değişkenler
- 'Eğer' tümceleri
- Basit ve içiçe geçmiş 'Eğer' tümceleri
- Lambda fonksiyonlarındaki şartlı ifadeler
- Devam eden döngüler
- Kes ve devam et

4. Ünite: Veri Analizi Paketleri

- NumPy
- Ndarray
- Alt simgeleme ve dilimleme
- İşlemler
- Pandas
- Veri Yapısı
- Veri Düzenleme
- Gruplandırma ve Birleştirme

Data Science with Python: Data Analysis and Visualization (2 gün)

1. Ünite: Python'a Giriş

Python üst düzey bir programlama dilidir. Python'daki temel söz dizimlerini ve veri yapılarını öğreneceksiniz. Kodları, etkileşimli ve keşifçi bir bilişim ortamı için güçlü ve verimli bir ortam sağlayan mükemmel bir araç olan Lpython Notebook'da gösteriyor ve çalıştırıyoruz.

- Lpython Notebook'a giriş
- Python'daki temel nesneler
- Değişkenler ve kendi kendini tanımlayan fonksiyonlar
- Kontrol akışı
- Veri yapıları

2. Ünite: Python ile Daha Derinleri Keşif

Python, nesne yönelimli bir programlama (OOP) dilidir. OOP konusunda temel seviyede bilgi sahibi olunması, Python kodlarının nasıl çalıştığını anlamanıza yardımcı olacaktır. Ayrıca kirli ve yapılandırılmamış verilerle de ilgilenmeniz gerekecektir. Düzenli ifadelerin uygulanması gibi verilerinizi temizlemenin bir sürü yolunu öğreneceksiniz.

- Nesne yönelimli programlamaya giriş
- Dosyalarla nasıl ilgilenilir
- Python komut metinlerini çalıştırma
- Dizileri ele alma ve işleme

3. Ünite: Bilimsel Bilişim Araçları

Python'u veri analizi için güçlü kılan iki bilimsel bilişim modülü vardır: Numpy ve Scipy. Numpy, Python'da bilimsel bilişim için temel pakettir. SciPy ise bilimsel bilişimi ele alan ve her geçen gün daha da genişleyen bir paketler topluluğudur.

- Numpy
- Scipy

4. Ünite: Veri Görselleştirme

Python, ayrıca "Matplotlib" ve "Seaborn" kullanarak grafik de oluşturabilir. Matplotlib, çizim ve diğer iki boyutlu veri görselleştirmelerinin üretilmesi için kullanılan en popüler Python kitaplığıdır. Seaborn ise Matplotlib'i esas alan bir Python görselleştirme kitaplığıdır. İstatistiksel grafiklerin çizilmesi için üst düzeyde bir arayüz sunar

- Seaborn
- Matplotlib

5. Ünite: Pandas ile veri manipülasyonu

Pandas, yapılandırılmış verilerle çalışmak için zengin veri yapıları ve fonksiyonları sunar. Pandas içindeki "DataFrame" nesnesi, R'deki "data.frame" nesnesi gibidir. Pandas da veri manipülasyonunu (filtreleme, seçme, gruplandırma, birleştirme, vs.) R'deki kadar kolay bir hale getirir.

- Pandas

3.Big Data with Hadoop and Spark (2 gün)

1. Ünite - Giriş: Hadoop, MapReduce, Python

- Büyük Veri ve Hadoop ekosistemi hakkında genel bilgi
- MapReduce kavramı
- HDFS – Hadoop Dağıtık Dosya Sistemi
- MapReduce için Python

2. Ünite – MapReduce

- MapReduce için daha fazla Python
- Python veri yayını ile MapReduce'un uygulanması

3. Ünite – Hive: Büyük Veri için veritabanı

- Hive kavramları, Hive sorgulama dili (HiveQL)
- Python'daki kullanıcı tanımlı fonksiyonlar (veri yayını kullanılarak)
- Python'dan Hive'a erişme

4. Ünite – Pig: MapReduce Kullanılarak Büyük Veri Kümelerini Analiz Platformu

- Apache Pig'e giriş
- Pig'de Veri Tipleri
- Pig Latin
- Pig'in MapReduce'da derlenmesi

5. Ünite – Spark

- PySpark kullanılarak Spark'a giriş
- Temel Spark kavramları: RDD'ler, dönüştürmeler, işlemler
- PairRDD'ler ve dönüştürmelerin birleştirilmesi
- Gelişmiş Spark: partisyonlar; ortak değişkenler
- SparkSQL

6. Ünite – Proje

- Örnek çalışmalar / Final projeleri