Data Analysis with Python

This course aims to familiarize the students with the most popular scripting language – Python and data science must know tools – Numpy,Pandas, Matplotlib and Seaborn. The course will start with learning the syntax of functional aspect of python and how you can develop scripts rapidly. It then goes to the data science must know tools and augmenting your python skills with shell scripting. This course aims to teach the fundamentals of data science. As a data scientist, you are not expected to write machine learning code everyday. A lot of work focus around gathering data, cleaning data and coming up with useful analysis and visualizing them. This course will prove to a strong base for students who want to take advance courses in data science.

Special Focus for Spring 2017 : Natural Language Processing

For Spring 2017, We will study how we can use NLP tools for Data Analysis. A large amount of data is available in form of human-readable text and analyzing them to gain insights in very valuable for any company. We will look at stages of NLP and some tools that will enable us to understand large amount of text data using open source tools like NLTK, numenta etc.

Syllabus :

Lecture	Topic
1	Introduction to Python: • How to move away from OOP and develop rapid scripts. • Numeric Type • String Fundamentals • Understanding Modules • Understanding the tools for development
2	Python Data Structures: • List, Dictionary and Tuples
3	Statement and Syntax: • Python Statements and assignments • Conditionals Constructs • Iterators
4	Functions and Generators • Basics of Functions • Variables Scopes and Arguments • Advance Function Topics • NLP : Introduction, Tokenization • Search Index
5	• Comprehensions and Generators • Exception Handling • NLP : POS Tagging • Introduction to NLTK
6	Numpy NLP : Using Parsers.
7	Midterm
8	Pandas • Introduction to pandas data structure • Data loading and storing • NLP : Name Entity Recognition
9	Advance Pandas concept - I • Cleaning, Transform and merging data. • Data Aggregation • NLP : Analyzing Documents - I

10	Advance Pandas concept – II • Data Grouping • Working with time series data • NLP : Analyzing Documents and extracting information
11	Matplotlib NLP : Chatbots using API.ai
12	Seaborn
13	Shell Scripting : • ZSH shell • AWK,SED commands • Cron Jobs.
14	Analyzing example data sets using NLP tools Create QA system bot.
15	Final Exam

Reference Books :

1. Learning Python, 5th Edition
   By Mark Lutz
   Publisher: O'Reilly Media

2. Python for Data Analysis
   By Wes McKinney Publisher: O'Reilly Media