



Introduction to Data Science

(Lecture 7)

Dr. Mohammad Pourhomayoun

Assistant Professor

Computer Science Department

California State University, Los Angeles



Data Science with Python

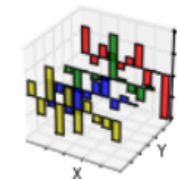
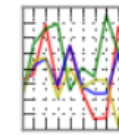
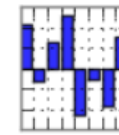
- After reviewing python syntax, now we can start working with Data Science libraries , Data Processing libraries , and Machine Learning libraries and packages in Python.



Data Science with Python

IP[y]: IPython
Interactive Computing

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



scikits
learn
machine learning in Python

NumPy

matplotlib

SciPy.org

Sponsored By
ENTHOUGHT



Numpy Package

Numpy

- **Numpy** provides a lot of powerful functions for advanced mathematical operations. It provides objects and data structures to process large, multi-dimensional arrays and matrices in very short time.
- You will learn some capabilities of Numpy in future tutorials.



Pandas Package

Pandas

- **Pandas** is a powerful package to read, manipulate, and process large-scale data.
- **Pandas** introduces two new data structures to Python, **Series** and **DataFrame**, both of which are built on top of NumPy n-dimensional array.
- A **Series** is a one-dimensional vector similar to an array, list, or column in a table. It will assign an index to each item in the Series (starting from 0).
- A **DataFrame** is a structured data table (or Matrix) comprised of **rows** and **columns**.
- Each column of a DataFrame can be considered as a Series.
- You will learn the properties and capabilities of Pandas in next tutorials.

IRIS Dataset

- **Recognizing flowers**

- 150 sample flowers in three species (50 each).
- Species of Iris (Labels): setosa, versicolor, virginica
- Features: sepal length, sepal width, petal length, petal width



IRIS Dataset

- We start with Iris dataset.
- More information about iris dataset:
<https://archive.ics.uci.edu/ml/datasets/Iris>
- To download the iris data:
<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>
- Since iris dataset is very popular and widely used, it is also already embedded in sklearn library.

Data Science Practical Tutorial

- Let's open file *CS4661-PythonDataScienceTutorial-Lab2.ipynb* in Jupyter notebook to continue the tutorial.

