Python Lecture Notes on Data Analysis
Scientific Programming, Machine Learning

Zime, Songbian PhD

September 18, 2023

Why Python?

- Introduction to the course
- The point of view of Web Mining

Introduction to Scientific Programming

- Programming Basics
  - Exercise 2.1: Python Standard Data Structures
- Numpy in a Nutshell
  - Exercise 3.1: Data Structures and Operations in Numpy
- Scipy in a Nutshell
- Matplotlib in a Nutshell
  - Exercise DSiP-5-1: Analyzing the Iris Dataset with Matplotlib
- Pandas-based Data Analysis
  - Exercise 6.1: Analyzing New York Open Data with Pandas

Machine Learning in a Nutshell with scikit-learn

- Overview and Preprocessing
- Supervised Learning: Classification and Regression
- Unsupervised Learning: Clustering
- Unsupervised Learning: Projections and Manifolds

Machine Learning Basics

- On the Data
- Regression
    - A simple Example
    - Least Mean Square Algorithm - a Gradient Descent approach to regression
    - Regression and Classification - using the analytical solution
- Concept Learning
- Measuring Performance

An Introduction to NLTK

- NLTK Exercise

Information Visualization with JavaScript and D3JS

Social Network Analysis Basics

Crawling and Analyzing Twitter

- Exercise: Crawling and Analyzing Twitter
- Project: Shitstorm Detection
- Project: Topic Detection
- Project: Web Crawling - Genre Classification