Data 690 - Data Structures and Algorithms in Python

Instructor

- Name: Donghwa Kim
- Email: donghwa.kim@umbc.edu
- Office Hours: Virtual by request.

Course Description

This course aims to provide an introduction of data structures and algorithm in Python. Students will also learn basic Python data science tools including NumPy, Pandas and Matplotlib. Data structure topics covered are arrays, strings, linked lists, stacks, queues, trees and graphs; Topics in algorithms include sorting and searching, divide-and-conquer and greedy strategies, dynamic programming, graph and compression algorithms using Python. The second half of the course will focus on three fundamental data science tools in Python - NumPy, Pandas and Matplotlib. The course will provide a broad introduction of Python programming and exposure to the data science tools. No prior programming experience is required.

Course Duration

- 08/31/2021 – 12/13/2021
- Meeting time: self-paced

Learning Outcomes

At the end of this course, students will

- Have introductory knowledge of Python programming (data types, functions, modules, object-oriented programming)
- Have familiarity of Python data science tools and libraries (Jupyter notebook, NumPy, Pandas, Matplotlib)
- Have knowledge of the most common Python abstractions for data collections (e.g., stacks, queues, lists, trees, maps).
- Understand algorithmic strategies for producing efficient realizations of common data structures.
- Be able to apply data structures and algorithms to solve complex problems.

Textbooks

J. V. Guttag, Introduction to Computation and Programming Using Python (2nd Ed.), MIT, 2016

J. VanderPlas, Python Data Science Handbook, O'Reilly, 2017 (Online: https://jakevdp.github.io/PythonDataScienceHandbook/index.html)

Tentative Schedule

Lecture	Subject	Details
1	Introduction	Week of 08/31/21 Introduction to Python - Installing Anaconda - Using Python interpreter - The basic elements of Python (Guttag, 2.1) - Branching Programs (Guttag, 2.2) - Strings and input (Guttag, 2.3) - Iteration (Guttag, 2.4) Introduction to Jupyter Notebook - Installing Jupyter notebook - Jupyter notebook basics Reading - Guttag, chapter 2 HW #1 Discussion #1
2	Python Functions	Week of 09/06/21 Simple Numerical Programs - Exhaustive Enumeration (Guttag, 3.1) - For loops (Guttag, 3.2) - Approximate solutions and bisection search (Guttag, 3.2) - Using floats (Guttag, 3.3) - Newton-Raphson (Guttag, 3.4) - Python lambda functions (https://realpython.com/python-lambda/) Functions, scoping and abstraction - Functions and Scoping (Guttag, 4.1) - Specifications (Guttag, 4.2) - Recursion (Guttag, 4.3) Reading - Guttag, chapters 3 & 4 HW #2 Quiz #1

3	Python Modules and Structured Types	Week of 09/13/21 Scoping, modules and files - Global Variables (Guttag, 4.4) - Modules (Guttag, 4.5) - Files (Guttag, 4.6) Structured Types, Mutability, and Higher-Order Functions - Tuples (Guttag, 5.1) - Ranges (Guttag, 5.2) - Lists and Mutability (Guttag, 5.3) - Functions and Objects (Guttag, 5.4) - String, Tuples, Ranges, and Lists (Guttag, 5.5) - Dictionaries (Guttag, 5.6) Reading - Guttag, chapters 4 and 5 HW #3 Quiz #2 Discussion #2
4	Python Testing and Exception Handling	Week of 09/20/21 Testing and Debugging - Testing (Guttag, 6.1) - Debugging (Guttag, 6.2) Exceptions and Assertions - Handling Exceptions (Guttag, 7.1) - Exceptions as a Control Flow Mechanism (Guttag 7.2) - Assertions (Guttag 7.3) Reading - Guttag, chapters 6 and 7 HW #4 Quiz #3
5	Python Object-Oriented Programming	Week of 09/27/21 Introduction to Object-Oriented Programming - Abstract Data Types and Classes (Guttag, 8.1) - Interitence (Guttag, 8.2) - Encapsulation and Information Hiding (Guttag, 8.3) Reading - Guttag, chapter 8 - https://www.youtube.com/watch?v=qiSCMNBIP2g - https://www.datacamp.com/community/tutorials/python-oop-tutorial

		HW #5 Quiz #4 Discussion #3
6	Algorithmic Complexity	Week of 10/04/21 Introduction to Algorithmic Complexity - Computational Complexity (Guttag, 9.1) - Asymptotic Notation (Guttag, 9.2) - Complexity Classes (Guttag, 9.3) Reading - Guttag, chapter 9 HW #6 Quiz #5
		Week of 03/14//21 Spring Break
7	Algorithms and Data Structures	Week of 10/11/21 Simple Algorithms and Data Structures - Search Algorithm (Guttag, 10.1) - Sorting Algorithm (Guttag, 10.2) - Hash Tables (Guttag, 10.3) - Advanced data structures:linked lists, stacks, queues and trees (Jupyter Notebook presentation) Reading - Guttag, chapter 10 - Jupyter notebook presentation for linked lists, stacks, queues and trees HW #7 Quiz #6 Discussion #4
8	Graph Optimization and Dynamic Programming	Week of 10/18/21 Knapsack and Graph Optimization - Knapsack Problems (Guttag, 12.1) - Graph Optimization Problems (Guttag, 12.2) Dynamic Programming - Fibonacci sequences, revisited (13.1) - 0/1 Knapsack problem (13.2) - Divide-and-conquer (13.3)

		Reading - Guttag, chapters 12 and 13 HW #8 Quiz #7
9	NumPy (1/2)	Week of 10/25/21 NumPy - Part 1 - Python Data Types review (https://jakevdp.github.io/PythonDataScienceHandbook/02.01-understanding-data-types.html) - Basics of NumPy Arrays (https://jakevdp.github.io/PythonDataScienceHandbook/02.02-the-basics-of-numpy-arrays.html) - Computation on NumPy arrays (https://jakevdp.github.io/PythonDataScienceHandbook/02.03-computation-on-arrays-ufuncs.html) - Aggregations (https://jakevdp.github.io/PythonDataScienceHandbook/02.04-computation-on-arrays-aggregates.html) - Broadcasting (https://jakevdp.github.io/PythonDataScienceHandbook/02.05-computation-on-arrays-broadcasting.html) Reading - VanderPlas, chapters 1 and 2 HW #9 Quiz #8
10	NumPy (2/2)	Week of 11/01/21 NumPy - Part 2 - Comparisons, Masks, and Boolean Logic (https://jakevdp.github.io/PythonDataScienceHandbook/02.06-boolean-arrays-and-masks.html) - Fancy indexing (https://jakevdp.github.io/PythonDataScienceHandbook/02.07-fancy-indexing.html) - Sorting Arrays (https://jakevdp.github.io/PythonDataScienceHandbook/02.08-sorting.html) - NumPy Structured Arrays (https://jakevdp.github.io/PythonDataScienceHandbook/02.09-structured-data-numpy.html) Reading - VanderPlas, chapter 2 HW #10 Quiz #9

11	Pandas - Introduction (1/3)	Week of 11/08/21 Pandas - Part 1 - Introducing Pandas Objects (https://jakevdp.github.io/PythonDataScienceHandbook/03.01-introducing-pandas-objects.html) - Data Indexing and Selection (https://jakevdp.github.io/PythonDataScienceHandbook/03.02-data-indexing-and-selection.html) - Operating on Data in Pandas (https://jakevdp.github.io/PythonDataScienceHandbook/03.03-operations-in-pandas.html) - Handling Missing Data (https://jakevdp.github.io/PythonDataScienceHandbook/03.04-missing-values.html) Reading - VanderPlas, chapter 3 HW #11 Quiz #10
12	Pandas (2/3)	Week of 11/15/21 Pandas - Part 2 - Hierarchical Indexing (https://jakevdp.github.io/PythonDataScienceHandbook/03.05-hierarchical-indexing.html) - Concat and Append (https://jakevdp.github.io/PythonDataScienceHandbook/03.06-concat-and-append.html0 - Merge and Join (https://jakevdp.github.io/PythonDataScienceHandbook/03.07-merge-and-join.html) - Aggregation and Grouping (https://jakevdp.github.io/PythonDataScienceHandbook/03.08-aggregation-and-grouping.html) Reading - VanderPlas, chapter 3 HW #12 Quiz #11
13	Pandas (3/3)	Week of 11/29/21 Pandas - Part 3 - Pivot Tables (https://jakevdp.github.io/PythonDataScienceHandbook/03.09-pivot-tables.html) - Vectorized String Operations (https://jakevdp.github.io/PythonDataScienceHandbook/03.10-working-with-strings.html)

		- Working with Time Series (https://jakevdp.github.io/PythonDataScienceHandbook/03.11-working-with-time-series.html) - High-Performance Pandas (https://jakevdp.github.io/PythonDataScienceHandbook/03.12-performance-eval-and-query.html) Reading - VanderPlas, chapter 3 HW #13 Quiz #12
14	Matplotlib - Introduction	Week of 12/06/21 Matplotlib - Simple Line Plots (https://jakevdp.github.io/PythonDataScienceHandbook/04.01-simple-line-plots.html) - Simple Scatter Plots (https://jakevdp.github.io/PythonDataScienceHandbook/04.02-simple-scatter-plots.html) - Visualizing Errors (https://jakevdp.github.io/PythonDataScienceHandbook/04.03-errorbars.html) - Density and Contour Plots (https://jakevdp.github.io/PythonDataScienceHandbook/04.04-density-and-contour-plots.html) - Histograms (https://jakevdp.github.io/PythonDataScienceHandbook/04.05-histograms-and-binnings.html) Reading - VanderPlas, chapter 4 HW #14 Quiz #13
	Final project	**Final Project due (12/19/21 11:59:00 PM)**

Grade Rubric
Final Project: 40 %
Home works: 30%
Quizzes: 25 %
Discussions: 5%

For each home work, you can get 3 points.

- There will be total of 14 assignments.
- Total possible points for HWs are 42 (3 x 12).

For each quiz, you can get 3 points.

- There will be total of 13 quizzes.
- Total possible points for quizzes are 39 (3 x 13).

For each discussion, you can get 3 points.

- There will be 4 discussions.
- Total possible points for discussions are 12 (3 x 4).

Home works, quizzes and discussions will be summed up and multiplied by respective weights

- Sum of home works x 30%
- Sum of quizzes x 25%
- Sum of discussions x 5%

Final project will have 30 total possible points. It will be multiplied by the weight of 40%.