MATH 183- Introduction to Data Science in Python

January 11, 2023

In this Introduction to Data Science in Python course, student will learn the basic concepts and skills in data science that are necessary for subsequent courses Data Science, such as Machine Learning. These concepts and skills include: expressing and storing real-world data using numbers, lists, dictionaries, arrays and tables; manipulating data; obtaining descriptive statistics and visualizing data. Students will also learn the basic python skills

- Create and assign variables
- Write programs with loops
- Write programs with conditions
- Define and use functions
- Work with basic dataset and data tasks

The following topics will be taught:

Python Basics

- Installation: Python, Jupyter notebook
- Interactive mode and script mode with basic operation
- Input/Outputs; If Statement; While and for Loops.
- Tuple, List, Set, Dictionary
- Function; File, Exception, Recursion
- Markdown and Scientific Writing in Latex.

NumPy

- NumPy I: Basics of Numpy array
- NumPy II: Computation on Numpy Array
- NumPy III: Aggregrations: Max, Min and things in bwtween
- NumPy IV: Broadcasting, Indexing, Sorting

Matplotlib

- Visualization with Matplotlib I: line plot and scatter plot
- Visualization with Matplotlib II: Histograms and Density

- Visualization with Matplotlib III: 3D plot, Multiple subplots, customizing legends and colorbars

Pandas DataFrame

- Data Manipulation with Pandas I: Indexing, Selection and Basic Operations
- Data Manipulation with Pandas II: Merge and Join
- Data Manipulation with Pandas III: Aggregation and Grouping

(if time permits) Data crawling and preprocessing, basic Machine Learning

- Data crawling
- Preprocessing
- Basic machine learning with scikit-learn

The material of the course will be largely from *Python Data Science Handbook—Essential tools for working with data*, by Jake VanderPlas and *Think Python: How to think like computer scientist* by Allen Downey. Notebooks for each lecture will be provided before each lecture.

Grading:

- **Homework (25%)** About 10 homework assignments due on weekly basis
- **Data Science Competition (15%)** As a special assignment, students take this class for credits are required to participate in the **visualization track** of the **2023 UI Data Science Competition**. If for some reasons, the competition does not take place, I will assign an alternative project/competition for this assignment.
- **In-class midterm Exam (25%)** The midterm exam will focus on your coding skills (for both algorithmic and data problems).
- **In-class final Exam and take-home project (40%)** The in-class exam will test your understanding of the course materials. The final project will give you an opportunity to work on a real world dateset of your choice.

The grades will be determined by:

- A: 90%-100%
- B: 80%-89.9%
- C: 70%-79.9%
- D: 60%-69.9%
- F: $\leq$ 59.9%

Learning Outcomes: The students will learn the basic concepts and skills in data science include: expressing and storing real-world data using numbers, lists, dictionaries, arrays and tables; manipulating data; obtaining descriptive statistics and visualizing data. The students will also learn the basic python skills to create an assign variables, to write programs with loops and conditions, and to construct and use functions. The students will also gain some experinces working with real world datasets.

Instructor: Dr. Linh Nguyen. Email: lnguyen@uidaho.edu.

Class time: 2:30-3:20 PM on MWF. Location: TLC 223.

Office Hours: Wednesday 11AM-12PM on WF (in person or on zoom 874 4002 1019, please email me in advance for zoom meetings).