AMOD 5310H
Data Science with Python
Summer 2020
Assignment 1

Due July 31(noon) via BlackBoard

Question 1 (30%)

For distance-based clustering algorithms, one of the time-consuming steps is to compute the distances. Perform some research online and summarize approaches that have been used by scientists to reduce the runtime of those clustering-based approaches (include references). Find at least two approaches and compare them to each other as well (roughly 500 words overall).

Question 2 (30%)

Use any scientific dataset of your choice that has at least 6 attributes and 1000 samples. Using matplotlib, plot the dataset using multiple views (small visualizations used for side-by-side comparisons). Discuss some relevant findings that you can detect using your chosen techniques. Also, explain how your chosen visualizations use preattentive features and how efficient they are encoding the attributes in your datasets. Hand in the documented source code and a roughly 500 word description of how you use the preattentive features to convey the findings to the viewer.

Question 4 (40%)

For this question, we are going to expand the image processing from the lab. You must use IDLE, and matplotlib for this question.

- a) Implement a Python program that either flips an image horizontally, or vertically, depending on user input
- b) implement a Python program that rotates a square image 90, 180 or 270 degrees clockwise, depending on user input

For each part, hand in your commented source code.