CSCE 5063 Machine Learning
Spring 2022

Mini Project Assignment 1
Date Assigned: Tuesday, Feb 1, 2022
Due Date: 11:59 PM, Thursday, February 10, 2022

The goal of this assignment is to:

1. Setup python, essential environment, and toolboxes useful for this course.
2. Make ourselves familiar with some of most common python toolboxes for machine learning and data science.

Task 1: Python Environment and toolbox setup

a) In your computers, set up python environment and some basic toolboxes that will be useful for you in machine learning and data science.
I have provided a **10-step instruction** for setup in course OneDrive link under MiniProject/Setup directory. If this is your first time setting up virtual environment for python this could be very helpful. The instruction is tested on Ubuntu 20. For MacOS user, it should work as well in most cases. If you have Windows system or Apple M1 MacBook, you could use similar approach and get help from internet resources to set up.

b) If you encountered any issue in setting up or have questions, feel free to post questions on Course Teams Link under Mini Project channel. Feel free to help each other out. I might not be able to help on this as each OS system/version could be different.

c) **Document and submit** (in Blackboard) all issues you encountered in setup process, provide your operating system information and how you finally solved the issue. This will help future students. If you did not encounter any issues and successfully setup and ran the SteupTest.ipynb code in first attempt, you do not need to submit anything in Blackboard for this part.

Task 2: The Fun part
During lecture 3, we introduced an amazing Python GitHub resource for the ISLR book.
https://github.com/JWarmenhoven/ISLR-python

Download the repository and play with Codes related to chapter 3. Try to compare the theory we studied in the class to the code implementation. If you are new to python or some of the toolboxes, we installed in Task 1, I recommend you spend at-least an hour reading brief introduction for each of the toolboxes in CSCE5063_tools.yml. You can use any internet resources

If you successfully setup your virtual environment and toolboxes as per the instruction given to you in Task1, you should be able to run all code cells of chapter 3 from the GitHub repository above.
If you **get error** relating to ".as_matrix", replace it with ".to_numpy".

You do not need to submit anything for Task 2. This is only for all of us to be familiar with some of toolboxes and start getting our hands dirty. I highly recommend all of you to go through code and play with.

Task 3: Multiple Linear Regression with interaction using *sklearn*.
In the GitHub repository above, Regression with Interaction between Radio and TV for advertisement data is done using smf toolbox (table 3.9).
Implement the same regression function using *sklearn* library.
Provide your code in the submission.
Find regression coeffs: Intercept, TV, Radio and Radio*TV. Are the values you calculated sklearn toolbox same as calculated using smf toolbox?

Submit all your submission into Blackboard in a single pdf file.