

Learning Report Summary (LRS)
by: Valerie Bixler
course: CIS2532-NET01
assignment: Hwk10 Machine Learning
due: Sun, 4/26/2020

Hello Dr. Shamsuddin,

I successfully completed homework assignment 10 for course CIS2532-NET01 which required learning about, reporting, and practicing with Machine Learning (ML). A research paper on ML helped me learn about the different types of Machine Learning, as did watching the videos indicated by the class instructions. The programs I wrote to practice a Seaborn pairplot graph and produce a simple linear regression helped me understand the power and abilities of machine learning.

I have tested the programs thoroughly and they run successfully with valid results. I had gone through much of the chapter, performing the exercises prior to working on the assignment. Doing the homework questions meant replaying much of that work with different datasets.

For the pairplot, I ran into a few issues. I had been able to easily get it working with the Iris dataset. For the California Housing dataset, I tried various solutions. I had recurring issues with density and KDE bandwidth. I was able to then get it working by paring down the code to simply doing the pairplot without parameters. I was able to produce the Matplotlib rendering, with the features of panning and zooming accessible via the icons in the lower left of the Matplotlib window. Actual panning and zooming moved frustratingly slowly though. I think it may be because I have a Windows computer with a cheap graphics card? There were also 20,640 plotpoints of California DataFrame versus the Iris DataFrame of 150 rows, 4 columns. I tried it with successively smaller California DataFrames and did see a slight improvement, but it is still slow to pan and zoom.

Programming the simple linear regression with the New York City temperatures from the NOAA's Climate at a Glance website was very interesting. It surprised me how few lines of code produced enlightening data visualization. Having to produce the January data by going to the www.ncdc.noaa.gov/cag website gave me an appreciation for this website and the data there too.

I spent more time on this program assignment than I think I should have. As you prefaced the assignment, machine learning is a big topic. In going through chapter 15 there is a lot of new terminology and concepts to cover. It was slow going. After working through most of the chapter, I had to switch over to the assignment to make sure I would have enough time to complete it. Even formatting the programs to confirm to Python best practices and adding comments to the simple linear regression program took me longer than I think it should have. I did all the work on this assignment myself,

though I did get help from my teenage son regarding the slope equation and its 'rise over run' calculation. I wanted a way to confirm mathematically that the visual representation was correct for the simple linear regression.

I appreciate this assignment for opening me up to the breadth, power, and data visualization functions available for Machine Learning.

Please let me know if you have any questions.