
EE / CPRE / SE 491 - Weekly Report 03

10/13/19 – 10/26/19

Group number: sdmay20-41
Project title: Machine Learning for Understanding Aging
Client &/Advisor: Dr. Julie Dickerson
Team Members/Role:

- Ian Simon / Chief Engineer
- Jacob Laing / Chief Engineer
- Nathan Carter / Test Engineer
- Samantha Williams / Meeting Scribe
- Scott Rose / Meeting Facilitator
- Aria Sheets / Report Manager

Weekly Summary:

- In the first week, we made progress in understanding DDI files, and how we should use it in our project. We made progress in finding a library to use in order to visualize our machine learn with the data (we chose Plotly due to its interactive graphs and models). We made progress in getting our own database server going by making a request to ISU and getting it approved. After we got it approved, we set up our credentials for the server.
- In the second week, we made progress in understanding different types of regressions: linear, ridge, and lasso. We compiled and ran code that would show the different types of regressions in different types of graphs.

Past Week Accomplishments:

- Ian Simon: Sami and I had trouble finding any resources on DDI files. We decided to contact the program director who oversees the MIDUS data at Michigan State. We sent him an email inquiring if they have any resources available or other data formats to collect the data from.
- Jacob Laing: Worked on researching different libraries that deal with linear regression within python and then also worked on ridge regression examples. I also got the vm setup with a database and added everyone to be able to access it.
- Nathan Carter: I worked on researching different libraries to visualize our data in python. After learning about a few different libraries our team came to the conclusion that we would use the plotly library. In the second week I worked with Aria on Lasso regression using the ski-kit library. Using a mix of sci-kit and matplotlib libraries we were able to perform lasso regression on a sample set.
- Samantha Williams: I researched ways to automate the reading of ddi files. There aren't any openly available on the web, but I did find some examples of github projects attempting to do this reading. We decided to try to email a contact from the organization that collects the MIDUS data to see if they have any tips for automating this process.

- Scott Rose: I worked with Jacob Laing to create an informational powerpoint on the linear regression model known as ridge regression. I also created a script and created experiments that would help my teammates become more familiar with the topic. Specifically, I created a program that showed how changing the alpha value on a ridge regression affected the predictions of a ridge regression model.
- Aria Sheets: I worked on finding a good library to plot our data within Python. There were many alternatives, but Plotly seemed like the best fit due to its interactive models. I also did work on understanding different types of regression techniques, primarily Lasso Regression. I worked on this with Nathan, and we got some code running that performed Lasso Regression on a diabetes dataset provided by sklearn.

Pending Issues:

- Ian Simon: Still waiting to hear back from the program director for the MIDUS data. May need to try another outlet to get the information from.
- Jacob Laing: Still working with ridge regression to really understand how it works and how we will be able to apply it to the MIDUS data.
- Nathan Carter: I am having an issue with determining how well a lasso regression line fits the data. Depending on what parameters I change I get different fits for a data set. I am not currently sure how to tell if one fit is better than another.
- Samantha Williams: Ian and I emailed the contact from the MIDUS data regarding reading ddi files, so we hope to hear from them soon. If we don't receive any tips, we will have to attempt to automate the reading on our own, or just manually read them. We also need to do our example of regression testing using the scilearn tutorial.
- Scott Rose: I will be trying to use a tool called Colectica Reader to try to read the DDI files. If the DDI files cannot be parsed by this application, I will continue to search for one that does.
- Aria Sheets: I don't have any pending issues as of now, but it took me much longer to understand how to use the sklearn library within Python so that I could use Lasso regression against the diabetes dataset.

Individual Contributions:

<u>Name</u>	<u>Individual Contributions</u>	<u>Hours this week</u>	<u>Hours Cumulative</u>
Ian Simon	Emailed the MIDUS program director for information on their DDI file types. Researched DDI files to try to find a solution for reading the MIDUS codebook information	Week 5: 3 Week 6: 6	22
Jacob Laing	Worked on researching different libraries that deal with linear regression within python and then also worked on ridge regression examples. I also got the vm setup with a database and added everyone to be able to access it.	Week 5: 4 Week 6: 5	22
Nathan Carter	Researched data visualization libraries in python. Continued researching neural networks and machine learning. Worked on lasso regression with aria using Sci-Kit learn and various other libraries.	Week 5: 4 Week 6: 6	24
Samantha Williams	Researched ways to automate the reading of ddi files. Attempted to parse the files. Contacted MIDUS contact.	Week 5: 2 Week 6: 6	23
Scott Rose	Created informational powerpoint on understanding ridge regression. Created a python script that demonstrated how effecting the alpha values of a ridge regression changes the results of that model.	Week 5: 4 Week 6: 4	21
Aria Sheets	Did work on trying to figure out what library we should use for plotting our information. Read through code using Plotly, and read about its advantages and disadvantages. Worked on understanding Lasso regression using a diabetes dataset by sklearn. Compiled code that showed Lasso regression in work against the diabetes dataset with Nathan.	Week 5: 4 Week 6: 6	25

Comments and Extended Discussion (Optional):

- Understanding the different regression techniques was fun to do, and it was interesting to see how the three different types (linear, ridge, and lasso) of regression are different.

Plans for the Upcoming Week:

- Ian Simon: I will be working on one of the regression types before our Tuesday meeting. Secondly we will have to follow up with the program director for MIDUS.
- Jacob Laing: Continue working with ridge regression and apply it to different data sets that more similarly represent what we are going to be using them for.
- Nathan Carter: I will continue work on lasso regression. I will try to focus more of my time on learning how to improve the lasso regression fit. I will also begin to work with plotly instead of matplotlib. I hope to get the lasso regression line and diabetes data set plotted with plotly.
- Samantha Williams: Next week I am going to finish the tutorial for regression in order to compare results with my teammates. After we hear back from the MIDUS contact, we can hopefully make more headway on reading those files.
- Scott Rose: Getting the Collectica reader to read the DDI files. Help with making improvements on the design document
- Aria Sheets: I plan to research more about regression, and how we can apply it to our project. I also plan on updating our design document with more details about how our team will function.