Random Lasso and Hi-Lasso Libraries

Spring 2019 Senior Project

Team Members

1. Sams Khan -**Student ID: 000538110** — skhan34@students.kennesaw.edu
2. James Matthew Hamilton -**Student ID: 000828373** — jhamil86@students.kennesaw.edu
3. Jason Wein -**Student ID: 000563745** — jwein2@students.kennesaw.edu
4. Seung Myeong Choo -**Student ID: 000449498** — schoo1@students.kennesaw.edu
5. Devyn Wilkins -**Student ID: 000586739** — dwilki22@students.kennesaw.edu

Project Outline

Regression analysis algorithms are becoming increasingly more fine-tuned for inferring gene expression in bioinformatics. Ridge regression was a robust starting ground for this series of regression analysis algorithms, but these have since improved in both accuracy and speed. Some examples of regression analysis that give a chronological scope of the progression this field are: lasso, net elastic, adaptive lasso, random lasso, and hi-lasso.
There is a lack of resources both educational and algorithmic on Random Lasso. Often researches will not attempt to implement Random Lasso due to its complexity, and opt for a less accurate but simple algorithm. Our goal is to release a highly malleable Random Lasso package for both Python and R. As an extension of this, we will also be releasing a package for Doctor Kangs new state-of-the-art Hi-Lasso algorithm.

Task Description

James Matthew Hamilton, Sams Khan

There exists no package for Random Lasso in R. There exists a lasso package called glmnet, but lasso is different than Random Lasso. Hi-Lasso is a new algorithm, so no packages exist for it as well. The goal is to submit this package to CRAN, so that a user can obtain it with a simple console command like install.packages(RandomLasso) and install.packages(Hi-Lasso) or a combination of both like install.packages(Lasso). The function will have several optional arguments for tuning the results and increasing the versatility of the function. The goal is for this function to be usable across as many data formats as possible.

1. Create a Package for Random Lasso in R
2. Create a Package for Hi-Lasso in R
3. Collaborate On Publishing Software Research Paper

Sams Khan, Jason Wein

Similar to R, there exists no package for Random Lasso in Python. The package goals for Python mirror the package goals for R. The R package and the Python package should have identical optional arguments. Because creating tasking for the R package and the Python package is similar, algorithms for different tuning arguments may be tackled then mimicked across the two languages.

1. Create a Package for Random Lasso in Python
2. Create a Package for Hi-Lasso in Python
3. Collaborate On Publishing Software Research Paper

Seung Myeong Choo, Devyn Wilkins, Jason Wein

Due to the lack of educational materials on Random Lasso and Hi-Lasso, a website demonstrating expertise in this field would benefit academia. This website will also serve as a tutorial for the R and Python packages. General information about the project and its contributors will also be on this webpage. A more ambitious goal is to implement some real-time data visualization tools/UI on the website using the Shiny package in R. This requires some knowledge of reactive variables. Much of the information published within the research paper will be mimicked on the website, so collaboration on the paper is also critical.

1. Create Documentation on Packages Usage for Website.
2. Create Educational Material on Random Lasso and Hi-Lasso for Website.
3. Create Project Description and Biography Page for Website.
4. Create a Data Visualization Demo for Website
5. Collaborate On Publishing Software Research Paper.

Tasking And Organization

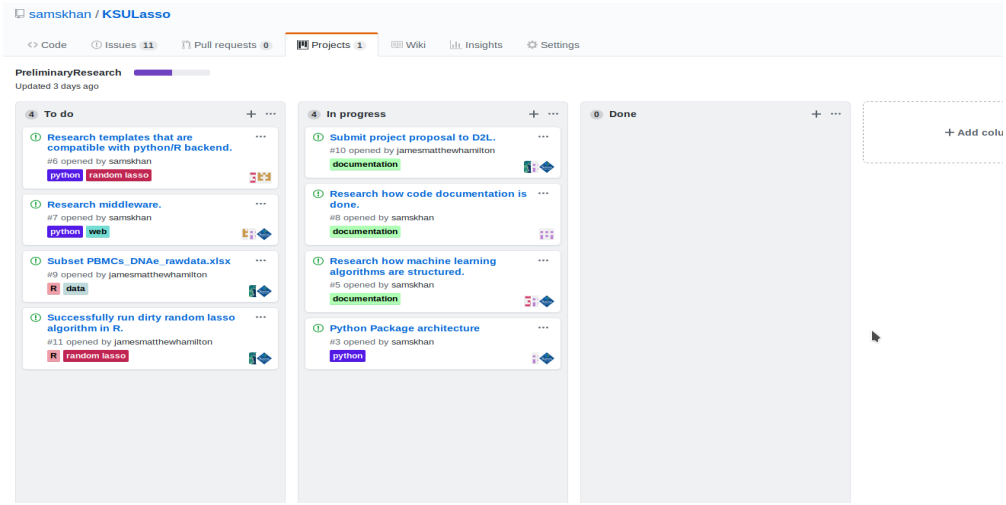

Figure 1: Week 1 Sprint and task allocation.

These tasking outlines do not mean we will strictly work on our parts, we plan on breaking down these bigger tasks into smaller tasks and assigning sub parts to interested team members. We want our members to know do some work on every aspect of the project. We are following the agile methodology using GitHub to post issues and resolutions in weekly sprints as shown in 1. We are currently in more of an exploratory phase. Multiple team members can be assigned to one task. Sams is currently our scrum master. Three submodual directories exist: (1) R, (2) Python, and (3) Web. Those with knowledge of git and heavier coding portions of the project can commit directly to the repository from their branch weekly. We communicate through Doctor Kang's DataX labs Slack channel.