Labs: Group 2 & 3 code snippets, **Assignment 5 Feedback** **Project Updates** **Check-in during the Lab** (One-on-One with the Instructor on Project Progress Updates)

Thilanka Munasinghe

Data Analytics

ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960

Group 4, Lab 7, October 25th , 2022

Project Update: National Oak Ridge Lab

- Student who are working on the projects introduced by National Oak Ridge Lab, please come today, 10/25/2022 at 2:15 pm ET to CII 3206 (HEALS Lab) discuss your project with scientists from National Oka Ridge Lab (Assaf Anyamba and Heidi). You can show your Assignment 5 Project Proposal Presentation and get the feedback. This meeting with them will help you to do the project effectively and their guidance on datasets and problem domain expertise will help you to chose and apply the correct models for your Final Project.
- You do NOT need to have the project progress results or models ready by today, just tell what problem you are working on and what datasets you are planning to use and how you will use them to build models and ask for any advice or guidance they can give you to do your project well. You can tell them what you have done so far using your Assignment 5 presentations slides.
- If you are planning to submit your work as a conference paper at the end of the semester, please tell them (I Highly recommend submitting to a conference), they will help you as how they have done in the past with our students. This is a great opportunity for you to learn from them and build a professional relationship with the scientists from a National Lab.
- Classroom Location: CII 3206 (HEALS Lab)
- Date: Today, 10/25/2022
- Time: 2:15 pm
- WebEx Meeting Link(if you can not come to CII 3206 class at 2:15 pm) :
- RPI-National Oak Ridge Lab: Data Analytics Project Update
- Tuesday, Oct 25, 2022 2:15 pm | 1 hour | (UTC-04:00) Eastern Time (US & Canada)

- **WebEx Meeting Link:**
- **https://rensselaer.webex.com/rensselaer/j.php?MTID=m0b6e91e3121ee7348e7944042f6c1c05**

Password: DataRPI2022

Today…

Project Check-in during the Lab:

- Assignment 5 (Project Proposal Feedback) during the
- One-on-One session with the Instructor project progress updates during the Labs.

You need to meet with instructor and show your current progress and the development of your term project (Assignment6).

Rpart – recursive partitioning and Conditional Inference

Reminder to go over these code snippets…

group3/lab1_rpart1.R

group3/lab1_rpart2.R

group3/lab1_rpart3.R

group3/lab1_rpart4.R

Try rpart for “Rings” on the Abalone dataset

group3/lab1_ctree1.R

group3/lab1_ctree2.R

group3/lab1_ctree3.R

Code snippets are available (script fragments in R available) on:
https://rpi.box.com/s/lu00cugurbk5mdvr0u58ztwsjufy57n2

Scripts – work through these

Next…

See in folder group2 and group3/ Labs

Go over the following scrips,

Lab3_ctree1.R

Lab3_ctree2.R

Lab3_ctree3.R

…..

And the remaining code snippets in group2/Lab 2 and Lab3

Search before you ask! You might need to search your code errors online when you are debugging your code!

script fragments in R available on:
https://rpi.box.com/s/2xx9ul1fmc6bf5ff8h4jreae69emikmf

NOTE: you are allowed to work in small groups and discuss during this lab.

Trees for the Titanic

Reminder to complete the following if you have not done yet…
After you complete this task, make sure to check-in with the TA
(Show your work to the TA and get-checked)

data(Titanic)

rpart, ctree, hclust for:

Survived ~ .

Reminder: Work through remaining code snippets in Group 2 & 3

These code snippets are available in the course repository:

(script fragments) in R available on:

https://rpi.box.com/s/2xx9ul1fmc6bf5ff8h4jreae69emikmf

Push your code to GitHub at the end of each lab, TA and I will check your code.

Next week: Support Vector Machine (SVM)

Please go over the reading material that available on our course webpage
for Class 7 prior to next week..

Class 7 Reading Assignment available at:
https://tw.rpi.edu/classes/data-analytics-fall-2022

Next week we will go over the SVM lecture,
please go over these articles a preparation for the lecture.

- https://rpi.box.com/s/umcnnuezeb7edh2qjw0vqf65w1n90hqg
- https://rpi.box.com/s/8cyw2gomptjb7u7u14hikxceqflv4q06
- https://rpi.box.com/s/o06ey08hhflvjjgqrkkncpdjincap3b8

Reading Assignment on R-SVM

- http://www.stanford.edu/group/wonglab/RSVMpage/R-SVM.html
  - Read/ skim the paper
  - Explore this method on a dataset of your choice, e.g. one of the R built-in datasets

ggplot - line graph example

Code snippet to practice ggplot- line graphs…
Make sure to go over the chapter 4 of gcookbook…

```
# Chapter4_Line_Graphs_R_Graphics
library(gcookbook)
ggplot(BOD, aes(x=Time, y=demand)) + geom_line()
BOD
BOD1 <- BOD # make a copy of the dataset
BOD1$Time <- factor(BOD1$Time)
ggplot(BOD1, aes(x=Time, y=demand, group=1)) + geom_line()
ggplot(BOD, aes(x=Time, y= demand)) +geom_line() + ylim(0, max(BOD$demand))
ggplot(BOD, aes(x=Time, y=demand)) +geom_line() + expand_limits(y=0)
# Adding points to a line graph
ggplot(BOD, aes(x=Time, y=demand)) + geom_line() + geom_point()
library(gcookbook) # For the data set
ggplot(worldpop, aes(x=Year, y=Population)) + geom_line() + geom_point()
# same with log-y axis
ggplot(worldpop, aes(x=Year, y=Population)) + geom_line() + geom_point() + scale_y_log10()
```

Today…

Assignment 5 (Proposal Presentation) Feedback

Project Check-in during the Lab:

Remaining One-on-One with the Instructor on Project Progress Updates during the Labs. You Must check-in with the instructor before end of the class.

Show what you have completed so far…(Work in progress)

- You need to meet with instructor and show your current progress and the development of your term project (Assignment6).
- Instructor One-on-One meeting link: https://rensselaer.webex.com/meet/munast