BUS41204 Review Session 1

Introduction to R

Jingyu He
jingyuhe@chicagobooth.edu

01/07/2017

Plan

- R, Rstudio
- rmarkdown
- How to do kNN and cross validation in R

What's R

An open-source language for statistical computing and graphics.

- It's popular, have lots of existing packages.
- User-friendly.
- Free!

Someone might ask : why not Python?

- Feel free to use any language you want. Python also has machine learning libraries.
- We recommend **R**. But we do not provide tech support for any language other than R.

R and Rstudio

- The original R interface is only a command line.
- Rstudio is a fancy interface of R (also free!).
  - See your workspace (variables, functions & data)
  - Write scripts
  - Manipulate files, manage plots
  - It helps a lot when you write a big project.
- You may use Rstudio on your own laptop or Booth's clusters.

Install R and Rstudio

- Download R: http://cran.r-project.org
- Download Rstudio: http://www.rstudio.com/products/rstudio/download
- Booth online Rstudio: http://rstudio.chicagobooth.edu
  - Do computation on remote cluster.
  - Must login with your Booth ID (NOT CNetID)
  - If you don't have a Booth ID, request a temporary one from helpdesk@chicagobooth.edu

R packages

R can do many statistical analysis. Functions are organized in "pacakges" or "libraries". People can make contributions to the community by wrapping their own code as packages.

Install the package

```
install.packages("package_name")
```

Load the package

```
library("package_name")
```

Necessary R packages for this course

Required Run the code below in your R terminal to install all necessary packages for this course. It takes about 10 minutes.

```
packageNames = c("MASS", "ISLR", "animation",
"ElemStatLearn", "glmnet", "textir", "nnet",
"methods", "statmod", "stats", "graphics",
"RCurl", "jsonlite", "tools", "utils",
"data.table", "gbm", "ggplot2", "randomForest",
"tree", "class", "kknn", "e1071",
"data.table", "recommenderlab")

for (pkgName in packageNames) {
if (!(pkgName %in% rownames(installed.packages()))) {
  install.packages(pkgName,
  dependencies=TRUE, repos='http://cran.rstudio.com')
}}

update.packages(ask=FALSE)
```

Set Directory

For each project, put everything under one folder.

```
setwd("path to your project folder")
```

Good R tutorials

See the course webpage
https://chicagoboothml.github.io/ML2016/computing/

rmarkdown

Rmarkdown generates high quality documents with R raw code and outputs. **Not required.** If you feel rmarkdown too complicated to use, feel free to use Microsoft Word.

- a package "rmarkdown" + Rstudio + latex
- Write in markdown language, the package can compile documents for you.
- This slides was written in this way.
- Writing your homework / project by markdown is strongly recommended.
- Find more info in https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet, http://chicagoboothml.github.io/MachineLearning_Fall2015/Tutorials/R%20Markdown%20Tutorial/
- A template https://raw.githubusercontent.com/ChicagoBoothML/ML2016/master/code/BostonHousing_KNN_BiasVarTradeOff_CrossValid.Rmd

Seeking help

- Read R documents

```
help(kknn)
help("any function name")
```

- Google it!
- **Ask questions on piazza**. TA will check out piazza regularly.