ST720 Data Science

Syllabus

Seung Jun Shin (sjshin@korea.ac.kr)

Department of Statistics, Korea University

August 30, 2019

What We Learn?

- Basics on Data Science (Visualization, Handling, Modeling)
- Text Mining
- Machine Learning Theory (Binary Classification, Dimension Reduction)
- and more ...

Course Level

- This course is designed for the graduate students majoring in Statistics.
- Students should be familiar with calculous, matrix algebra, probability, mathematical statistics, and `R` programing.
- The first half (Data Visualization Transformation, i.e., EDA) is mostly about programing, while the latter half (Modeling and Machine learning) is quite technical and involves a lot statistical theories.

Homework

- Homeworks is to be assigned irregularly.
- Due date is a week after each homework assigned.
- Students must submit the homework BEFORE the class.
- Extensions will NOT be granted and any assignment not submitted on time will receive a zero.

Evaluation

- Attendance.
- Assignments.
- Midterm: Takehome.
- Final: Individual capstone project.

Reference

- Wickham and Grolemund (2017) R for Data Science.
- Wickham (2016) ggplot2: elegant graphics for data analysis, 2nd Edition.
- Silge and Robinson (2017) Text Mining with R: A tidy approach.
- Friedman, Hastie, and Tibshirani (2009) The elements of statistical learning: data mining, inference and prediction.
- and related articles....

Software

- R (www.r-project.org)
- Editor: R-studio (www.rstudio.com)