CAP 5768 Introduction to Data Science

Credits: 3 credits

Textbook, title, author, and year: R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, by Hadley Wickham and Garrett Grolemund. O'Reilly, 2017, ISBN-13: 978-1491910399.

** Free ebook from author website https://r4ds.had.co.nz/

An Introduction to Statistical Learning: with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Springer, 2017, ISBN-13: 978-1-4614-7137-0.

** Free ebook from author website http://faculty.marshall.usc.edu/gareth-james/ISL/

Reference materials: None

Specific course information

Catalog description: This course provides a comprehensive introduction to the tools and analysis workflows employed by data   scientists that include data wrangling, visualization, exploration, and modeling. Specific topics include an   overview of the field of data science and analytics, data visualization, exploratory data analysis, data transformation, parameter estimation, hypothesis testing, linear regression analysis, logistic regression classification, model selection, feature selection, dimensionality reduction, and clustering. The practical application of these techniques to real data, as well as the interpretation and presentation of analysis results, will be emphasized throughout the course.

Prerequisites: Graduate standing or permission of instructor

Specific goals for the course:
- Apply the R software suite and its associated packages to perform an array of data analysis techniques.
- Effectively manipulate, curate, visualize, and explore data and draw conclusions from this data.
- Identify appropriate statistical models to address diverse problems in data analytics.

Brief list of topics to be covered:
- Introduction to data science
- Introduction to R and data visualization
- Data transformations
- Exploratory data analysis
- Linear regression
- Model selection, feature selection, and regularization
- Classification with logistic regression
- Unsupervised learning