CSX415
Data Science Principals and Practice

Christopher Brown

U.C. Berkeley / Decision Patterns LLC

Spring 2018

Class Overview

Prerequisites

- R Programming
- Practical Machine Learning with R

Expectation

Install, maintain, R & Rstudio

Install R packages from CRAN and Github

Code in R

Train, create, evaluate ML models

Create R Markdown documents

Collaborate using Git in social workflow work flow

Always bring your laptop to class

Class / Objectives

- Engage Business Units

- Frame problems to make the suitable for solution via data science techniques.

- Cost and Scope Data Science Projects

- Efficiently solve problems

- Collaborate in a group using tools for collaborative/social programming

- Solve Problems

- Deploy solutions to operations

- Generate high quality, graphical and textual results

- Maintenance solutions

Berkeley
UNIVERSITY OF CALIFORNIA

Class Text: None

Supplemental Text

Applied Predictive Modeling
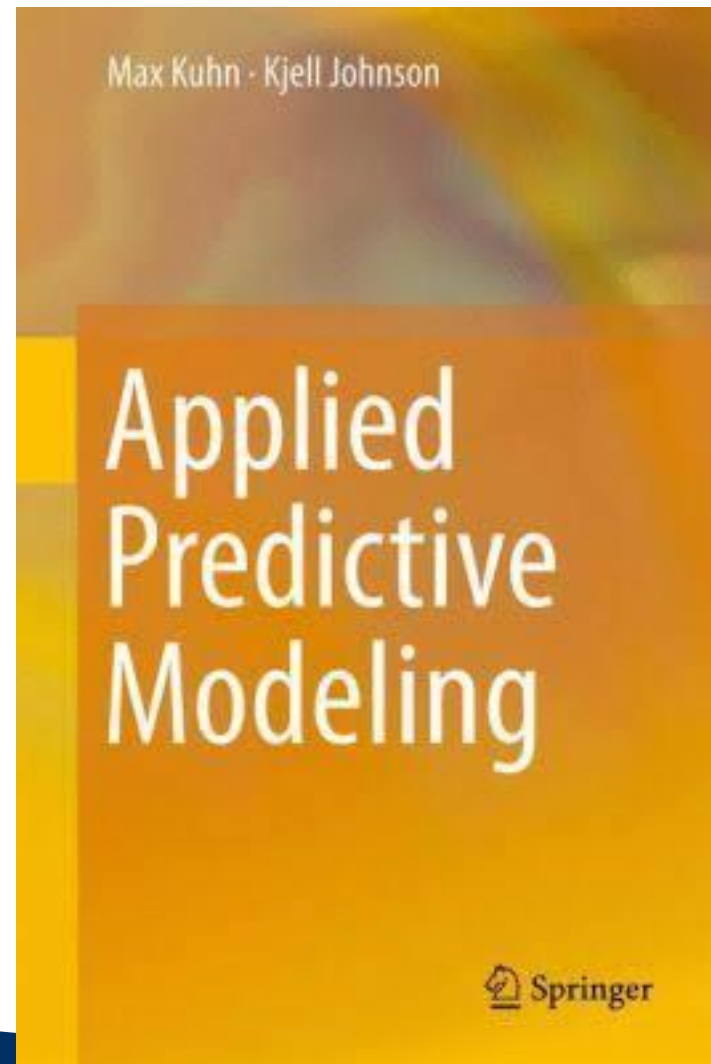
ISBN-13: 978-1461468486

ISBN-10: 1461468485

Kuhn, Max and Johnson, Kjell

Springer Science+Business

2013

KNOW YOUR DATA
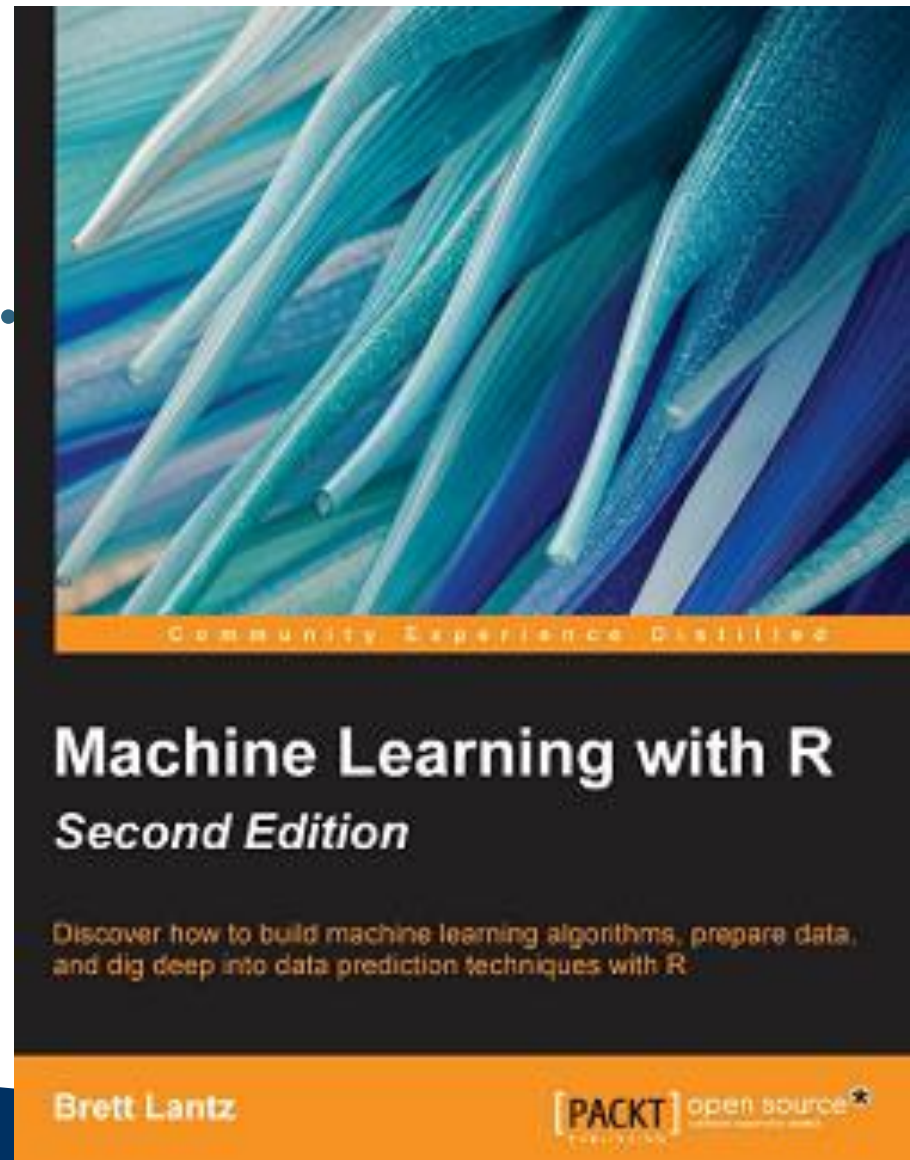
Supplemental Text

Machine Learning with R, 2nd Edition

ISBN: 978-1-78439-390-8

Lantz, Brett

Packt Publishing

2015

Additional Resources
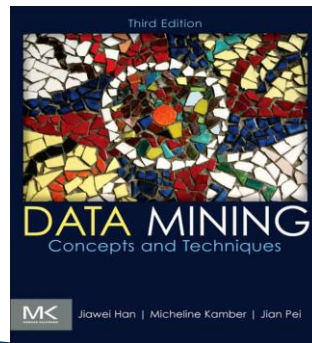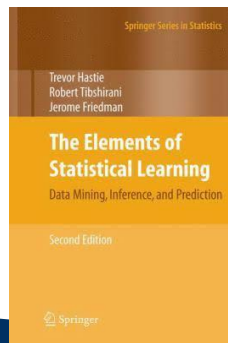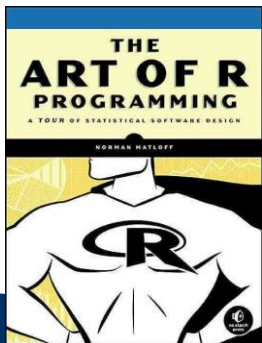
Texts
(not used in this class)

The Art of R Programming
  by Norm Matloff

Elements of Statistical Learning
  by Hastie, Friedman, Tibshirani

Data Mining Concepts and Techniques
  by Han, Kamber, Pei

Online

- CRAN
  - [Packages](#)
  - [Task Views](#)

- [Metacran](#) (r-pkg.org)

- [Stackoverflow.com](#)

- [r-bloggers.com](#)

- [H. Wickham Online Resources:](#)
  - [Advanced R Programming](#)
  - [R for Data Science](#)

- [Github](#)

Berkeley
UNIVERSITY OF CALIFORNIA

Contacts / Coordinates

- Christopher Brown
  christopher.brown@berkeley.edu
  checked once / day (mornings before 9AM)

- Class assignments: https://github.com/csx415

- Discussion Group TBD

Grading

- **5 Exercises (20%)**
  - Exercises are **Rmarkdown** in the github
  - Due at the Tuesday by 6:30 OM
  - Submitted via **github**
    - Please
    - ! Github commits are timestamped
  - Answers reviewed in class
  - Work on them in class, time-permitting
- **Quizzes (20%)**

- **Class Participation (20%)**

- **Project (40%)**

Berkeley
UNIVERSITY OF CALIFORNIA

Class Project

- Select an appropriate Data Science / Machine Learning problem

-
  Can be for at work and follow it from conception to inception.

Assignment

- Configure your laptop
- R
- Rstudio
- Rmarkdown
- Knitr
- Git, Github

- send me a short bio, github user id, and projects proposal

Attendance is mandatory

- No unexcused absenses