Advanced Topics in Data Science

Dr. Robin Donatello

MATH 499, Spring 2018

COURSE INFORMATION

- **Meeting time and locations:** TR 9:30-10:45am Holt 155
- **Prerequisites:**
  1. Introduction to Data Science (last offered as MATH 398 or CSCI 398).
  2. Math 456 (*or corequisite*)
  3. Or instructor permission

Course Description

This course aims to help students become successful Data Scientists by providing guidance and training in both soft, and advanced technical skills.

- Soft skills: Team collaboration tools, workflow automation. Dissemination of transparent research. Building a professional online presence. Ethics of predictive analytics, protecting privacy in the age of open data. Keeping current with a fast moving field. Critical thinking like a Data Scientist.
- Technical skills: Identifying and applying appropriate methods to answer a business or research need. Introduction to advanced statistical analysis techniques (Sample topics include: Non-Parametric techniques, Bayesian methods, Simulation methods, Network analysis, Experimental Design, Predictive modeling using Supervised and Unsupervised Machine Learning techniques, Modeling and mapping of Geospatial data). Introduction to current scalable technologies to handle Big Data.

Outline of topics

- Keeping up with the current state of Data Science
- Version Control
- Collaboration tools and Social media for data scientists
- Data analytics lifecycle
  - Data cleaning, modeling, predictions, dissemination , automation
- Statistical learning / machine learning methods
  - Prediction and classification
  - Tree based methods
  - Resampling methods
- Ethics and concerns around machine learning algorithms
- Big data
- Parallel processing, batch jobs

Materials

- **Computer**
  - A laptop is nearly invaluable, bring to class every day.
  - Loaner laptops will be available during class time only.
  - A solid internet connection is a must. We will be using a lot of online resources.

- **Textbooks/Readings**
  - R for Data Science http://r4ds.had.co.nz/ (free)
  - Happy Git and GitHub for the useR http://happygitwithr.com/ (free)
  - Introduction to Statistical Learning http://www-bcf.usc.edu/~gareth/ISL/ (free)
  - Weapons of Math Destruction, Cathy O'Neil (ISBN-10 0553418815) - Available for checkout in the library.
  - Various Data Science Blogs
- **Computer Software**
  - `R` http://cran.r-project.org/
  - `R-Studio` http://www.rstudio.com/products/rstudio/download/.
  - `LaTeX` (proTeXt or MacTex) https://www.latex-project.org/get/
- **Accounts**
  - GitHub https://github.com/
  - Slack https://slack.com/

Grading

The *approximate* contributions per category are: Discussions 15%, Peer review 15%, Assignments 35%, and Projects 35%.

Letter Grade Conversion: 100-90%: A, 89-80%: B, 79-70%: C, 69-60%: D, 0-59%: F

Americans with Disabilities Act

If you need course adaptations or accommodations because of a disability or chronic illness, or if you need to make special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. Please also contact Accessibility Resource Center (ARC) as they are the designated department responsible for approving and coordinating reasonable accommodations and services for students with disabilities. ARC will help you understand your rights and responsibilities under the Americans with Disabilities Act and provide you further assistance with requesting and arranging accommodations.

Accessibility Resource Center
530-898-5959
Student Services Center 170
arcdept@csuchico.edu

Last updated: Sun Jan 21 2018 3:07:42 PM