Exploratory Data Analysis & Inference: Big Data

Dr Jelena Bradic | MATH 189

Course Description

- This course is core class for the data science majors and aims to develop complete skills of data analysis, interpretation, visualization and hypothesis generation, etc.

- Pepper Canyon Hall 109
- Lectures: Tu/Thu at 6:30 pm
- Labs: Fridays at 5:00 pm -9pm
- Prerequisites: math 183 or similar
- Credits: 4

Course Objectives

Course Objectives	Results/Outcome Expected	Skills Developed
Make connections between different statistical concepts in unity	Able to make new scientific discoveries driven by data	Identification of appropriate statistical test, estimators, graphs and understand how they combine together to form a story
Formulate your own questions (not conclusions) from a real dataset	Able to design your own study	Formulate a set of questions/ hypothesis that together form a data-story
Scientific Method	Able to build up from existing work	Understand literature in a scientific area where data originated from
Coding, cleaning, planning	Able to design your own code: R or Python	Algorithms and coding strategies

Required Materials

Reading

- Open Intro to Statistics, https://www.openintro.org
- Professor Bradic's Lecture Slides
- Occasional reading materials (pdf files)
- In class blackboard work

Homework

- Approx Bi-Weekly assignments
- Discussion sections: R/Python Labs that supplement lectures and reinforce concepts

Technology

- Statistical programming language R, Python, ...
- Piazza

Instructional Methods

- Instructional methods are based on Project based and modeling based learning.
  - Lectures
  - Independent Reading (emphasized)
  - Class discussion/Piazza discussions
  - Self-driven homework completion/construction
  - Discussion sections (guided model based, guided examples)
    - topics will be introduced in addition to the ones done in lecture

Schedule

Week	Topic	Assignment/Project	Objective
Week 1	Introduction to Data		form groups for HW
Week 2	Case study I	linking smoking to survival of children	q-q plots, kurtosis, normal approximation
Week 3	Case study I		nonparametric testing using monte carlo
Week 4	Case study II	designing class-labs through video-games	survey data, bootstrap
Week 5	Case study II		classification trees, finite-sample correction

Schedule

Week	Topic	Assignment/Project	Objective
Week 6	Case Study III	searching for palindromes	Poisson process
Week 7	Case Study III, Final Project Choice		goodness of fit
Week 8	Case Study IV, Final Project Choice	snow gage and environment	longitudinal data
Week 9	Case Study IV		prediction intervals, regression
Week 10	Advanced topics	Causality, Deep nets	understand that there are more topics to learn

Assessment Criteria

- Bi-Weekly assignments, 45%
  - Group projects: 5-6 students
  - Written report: gradescope
  - Submitted Code: gradescope
  - Evaluations
    - Composition and presentation
    - Graphs and tables
    - Basic analysis
    - Advanced(independent) analysis
    - Technical evaluations
- Class participation, 5%
- Attendance: random checks,5%

- Final Project, 45%
  - Kaggle/Dream data competition datasets
  - Needs at least 2 weeks of work with 3 weeks leading to better results
  - Prof. will present 2-3 options for the choice of the dataset and each group will choose one project to work on

Resources

Web & Software Tools

- www.math189.edublogs.org
- Piazza
- R Studio or Keras

Office Hours

- Dr Bradic:
  Tuesday: 9:00-10:00am
- Denize Rava
  TBA
- Yuqian Zhang
  TBA

Labs

- Discussions: Friday
  6:00pm-9:50pm

Prof. Bradic

Email:
jbradic@ucsd.edu

Office:
AP&M 5151

Piazza private messaging

www.jelenabradic.net

Questions?