ER 190C: Data, Environment and Society

Instructor: Duncan Callaway, dcal@berkeley.edu
GSI: Seigi Karasaki, seigi@berkeley.edu

Fall, 2018

	Units: 4.0
Lecture: 102 Wheeler T/Th 9:30-11:00,	Lecture Hours per week: 3.0
Lab: 110 Barrows M 10-12	Lab Hours per week: 2.0
Office hours:	Barrows 399, Mon 12-1 (GSI)
	Barrows 310, We 3-4 by appt (intstructor)
	Barrows 310, Th 11:15-12:15 (instructor)

Course Description

This course will teach students to build, estimate and interpret models that describe phenomena in the broad area of energy and environmental decision-making. The effort will be divided between (i) learning a suite of data-driven modeling approaches, (ii) building the programming and computing tools to use those models and (iii) developing the expertise to formulate questions that are appropriate for available data and models. My goal is that students will leave the course as both critical *consumers* and responsible *producers* of data driven analysis.

We will work in Python in this course, and students must have taken Data 8 before enrolling. The course is designed to fit into Berkeley's emerging "data science" curriculum by providing students with a skill set similar to those developed in Data 100. However, in contrast to Data 100, here we will place a stronger emphasis on how to use prediction methods as decision-making tools in energy and environment contexts and less emphasis on web technologies, working with text, databases and statistical inference.

Materials

- You will need your own computer, but virtually any operating system will do (OSX, Windows, Linux, Chromebook).
- We will draw some material from Berkeley's Data 100 course book, freely available here: https://www.textbook.ds100.org
- Finally, we will draw material from the excellent text book, Introduction to Statistical Learning, available in both print and pdf form.

Prerequisites

Prerequisites:

- (required) Foundations of Data Science (CS/ INFO/ STAT C8)
- (recommended) Computing: An introductory programming course (CS61A or CS88).
- Math:
  - (required) High school or college calculus.
  - (recommended) Linear Algebra (Math 54, EE 16a, or Stat89a).

Course Structure

Class Structure

This is a four unit course, with three hours of lecture and two hours of lab section each week. Lectures will focus on theoretical and conceptual material but also introduce the programming structures required to use the material. Labs will be computer working sessions with a GSI and lab helpers available to work through weekly lab exercises.

Assessment

The course will have weekly labs and homework assignments, a mid term and a final project. Grading will be as follows:

- Homework: 20%
  - There will be eleven. We drop the lowest grade.
  - HW will be released on lab days (Monday) and due 8 days later, on Tuesday.
- Lab assignments: 20%
  - There will be eleven. We drop the lowest grade.
  - Released on Mondays (in lab) and due the subsequent Friday.
  - Attendance is 40% of lab grade, completing the lab is 60% of the grade.
  - Grading will focus on completeness rather than correctness.
- Mid-term: 25% (October 18, in class)
- Final project: 25% (due December 11 at midnight)
- Participation: 10% (Participation will be measured by answering questions in lecture with an online form.)

Late policy: You may request two extensions of two days over the course of the semester. Otherwise, we will note accept late homeworks and labs. Coordinate extension requests with the GSI.

Schedule and weekly learning goals

Lecture	Day	Topic	Methods Reading	Domain Reading	Homework
Lecture 1	2018-08-23	Course introduction			
Lab 1	2018-08-27	Getting started			HW1 - getting started
Lecture 2	2018-08-28	Data design		Science and Data Science	
Lecture 3	2018-08-30	Pandas, variable types (dict, tuple, etc), file types (json, xml, csv)	DS100 Ch3		
Lab 2	2018-09-03	Answer HW1 questions; Pandas			HW2 - pandas
Lecture 4	2018-09-04	Pandas, ctd, and energy and development.		Groh et al "What are we Measuring"	
Lecture 5	2018-09-06	Data cleaning and EDA	DS100 Ch4 and 5		
Lab 3	2018-09-10	Answer HW questions; EDA			HW3 - EDA
Lecture 6	2018-09-11	EDA and visualization		Goldemberg et al; Covers HDI, other energy metrics	
Lecture 7	2018-09-13	Visualization and data transformations	DS100 Ch6		
Lab 4	2018-09-17	Answer HW questions, visualization			HW4 - visualization
Lecture 8	2018-09-18	Introduction to modeling, KNN, Joshua Apte guest lecture		Lee, Miguel, Wolfram et al	
Lecture 9	2018-09-20	Modeling and estimation	DS100 Ch 10; ISLR Ch 2		
Lab 5	2018-09-24	Basic modeling, KNN			HW5: Basic modeling and linear regression
Lecture 10	2018-09-25	Linear Regression Part 1	ISLR 3.1	Alstone et al	

Continued on next page

Schedule and weekly learning goals

Lecture	Day	Topic	Methods Reading	Domain Reading	Homework
Lecture 11	2018-09-27	Multiple Linear Regression	ISLR 3.2		
Lab 6	2018-10-01	Answer HW questions, regularization			HW 6: Multiple linear regression, basic model selection, nonlinear features.
Lecture 12	2018-10-02	Extensions of the basic linear regression model.	ISLR 3.3	Marshall et al	
Lecture 13	2018-10-04	Gradient Descent	DS100 Ch 11		
Lab 7	2018-10-08	Ans HW questions; gradient descent			HW 7: Gradient descent; a "theory" homework.
Lecture 14	2018-10-09	Classification	ISLR 4.1-4.3		
Lecture 15	2018-10-11	Resampling	ISLR 5.1-5.2		
Lab 8	2018-10-15	Exam Review, through HW7 / Lecture 13.			
Lecture 16	2018-10-16	Exam Review, through HW7 / Lecture 13.			
Lecture 17	2018-10-18	Exam, through HW7 / Lecture 13.			
Lab 9	2018-10-22	Review classification and resampling			HW 8 - Resampling; A "theory" homework
Lecture 18	2018-10-23	Model Selection	ISLR 6.1		
Lecture 19	2018-10-25	Regularization and Shrinkage methods	ISLR 6.2		
Lab 10	2018-10-29				HW9 - Model selection with LUR data set
Lecture 20	2018-10-30	Classification and regression trees	ISLR 8.1		
Lecture 21	2018-11-01	Bagging, Random Forests, Boosting	ISLR 8.2		

Continued on next page

Schedule and weekly learning goals

Lecture	Day	Topic	Methods Reading	Domain Reading	Homework
Lab 11	2018-11-05				HW10 - Classification and regression trees with CES data
Lecture 22	2018-11-06	Wrap up tree methods			
Lecture 23	2018-11-08	Support vector machines	ISLR 9.1-9.2		
Lab 12	2018-11-12				HW11 – Support vector machines with CES data
Lecture 24	2018-11-13	Support vector machines, ctd	ISLR 9.3		
Lecture 25	2018-11-15	Wrap up support vector machines			
Lab 13	2018-11-19				
Lecture 26	2018-11-20	Neural networks	TBD		
OFF1	2018-11-22	Thanksgiving			
Lab 14	2018-11-26				
Lecture 27	2018-11-27	Wrap up neural networks	TBD		
OFF2	2018-11-29	Duncan travelling			NaN