U C L A

Computer Science Department

CS 188 **Data Science Fundamentals** **Winter 2020**

Instructor: Majid Sarrafzadeh
majid@cs.ucla.edu

CLASS TIME: MW 8 – 9:30 am
OFFICE HOURS: MW 9:30 to 10:30 am VI Room 393
PLACE: Kaplam A51
BOOK: **Doing Data Science: Straight Talk from the Frontline**
by Cathy O'Neil and Rachel Schutt
ISBN number: 978-1449358655

CLASS PAGE: **CCLE**

TA Office Hours (3256 S Boelter Hall or as otherwise noted) – All TA sections are on Fridays

TA	Sct.	Discussion	Office Hours	Email
Davis	1A	10am Rolfe Hall 3126	Mondays 1:30-3:30 PM, Boelter 3256S	tylerdavis@ucla.edu
Sajad	1B	12pm Kaplan Hall 169	Fridays 8-10AM Eng VI Rm. 464	sajad10@ucla.edu
Levine	1C	2pm Public Affairs Building 2238	Tuesdays 9-11am Engineering VI in front of room 391.	lionel@cs.ucla.edu

188 SPECIAL COURSES N LEC 001 MW 8:00A - 9:50A Kaplan A51 320 SARRAFZADEH, M. 120 15 01
Data Science Fundamentals
RESTRICTION: JUNIOR&SENIOR COM SCI, COM SCI&ENGR AND COM ENGR MAJORS

DIS 001A F 10:00A - 11:50A ROLFE 3126 52 TA 40 5
DIS 001B F 4:00P - 5:50P DODD 175 98 TA 40 5
DIS 001C F 2:00P - 3:50P PUB AFF 2238 46 TA 40 5

GRADING:

Homework	15%	3 homework assignments
Late MIDTERM	30%	8th week of the class (in class; 90 minutes)
Projects (3)	40%	3 projects
Class Participation	15%	

*Late policy: No late projects / homeworks will be accepted

DESCRIPTION:

A fundamental question that will be addresses is: given data arising in real-world, how does one analyze that data so as to understand the corresponding phenomenon. The course teaches critical concepts and skills in computer programming related to statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, health data, geographical data, and social networks.

We will cover topics in machine learning and data analytics. We learn about the two basic kinds of statistical models, which have classically been used for prediction. We also cover clustering methodologies. We then cover Feature Selection, feature Engineering, and Data Pipelines. We explore more sophisticated model evaluation approaches (cross-validation and bootstrapping) with the goal of understanding how we can make our models as generalizable as possible.

Sections:

1. Introduction: What Is Data Science? Modeling.
2. Statistical Inference, Exploratory Data Analysis, and the Data Science Process
3. Machine Learning Algorithms
4. Spam Filters, Naive Bayes, and Wrangling
5. Logistic Regression
6. Time Stamps and Financial Modeling.
7. Extracting Meaning from Data
8. Recommendation Engines: Building a User-Facing Data Product at Scale.
9. Data Visualization
10. Causality
11. Data Engineering

Prerequisites: CS 31, 32, 33 and a course in probability

Assignment	Assigned	Due
Project 1	Jan 8th	Jan 22nd
Hwk 1	Jan 15th	Jan 24th
Project 2	Jan 22nd	Feb 12th
Hwk 2	Jan 29th	Feb 14th
Project 3	Feb 12th	March 4th
Hwk 3	Feb 19th	March 6th
Midterm	Feb 26th	One hour: in class